

IDENTIFICATION DES GÈNES FACULTATIFS CHEZ *ESCHERICHIA COLI* À L'AIDE
DE MUTAGÉNÈSES ALÉATOIRES

Par

Frédéric Grenier

Mémoire présenté au Département de biologie en vue
de l'obtention du grade de maître ès sciences (M. Sc.)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, Septembre 2017

Le 21 décembre 2017

le jury a accepté le mémoire de Monsieur Frédéric Grenier dans sa version finale.

Membres du jury

Professeur Sébastien Rodrigue

Directeur de recherche

Département de Biologie

Professeur Pierre-Étienne Jacques

Codirecteur de recherche

Département de Biologie

Professeur Vincent Burrus

Évaluateur interne

Département de Biologie

Professeur Pascale Beauregard

Président-rapporteur

Département de Biologie

SOMMAIRE

La biologie synthétique est une discipline émergente consistant principalement en la modification génomique d'organismes vivants. Ces modifications complexes visent la programmation de systèmes biologiques pour accomplir des fonctions spécifiques parfois absentes dans la nature. Ce domaine se divise en plusieurs branches dont l'élaboration d'un châssis cellulaire minimal. En effet, depuis déjà plus d'une dizaine d'années, un grand intérêt est porté vers la création d'une cellule simplifiée. Le génome d'une telle cellule contiendrait seulement ce qui est nécessaire à sa survie en condition de laboratoire. Cette cellule posséderait pour principal avantage de permettre une meilleure compréhension de son fonctionnement global en plus d'éventuellement devenir un modèle d'étude pour la modification et la restructuration des génomes. Elle sera également plus fiable dans un contexte d'ingénierie métabolique puisqu'elle contiendra moins de circuits géniques, diminuant le nombre d'interactions non désirées. Une cellule minimale serait donc un sujet d'étude fascinant en plus d'être un organisme facilement reprogrammable, constituant ainsi une pierre angulaire pour le développement de la biologie synthétique.

La présente étude concerne la réduction génomique de la bactérie *Escherichia coli*. Nous avons choisi cette dernière puisque de nombreux outils moléculaires permettant de modifier son génome sont disponibles et puisque'une littérature très dense la concernant a été cumulée au fil des ans. Plusieurs projets de réduction ont été entamés avec cet organisme et les génomes produits ne semblent pas pouvoir atteindre moins de 3,0 Mpb sans présenter des défauts considérables de croissance. Nous avons donc sélectionné la souche simplifiée DGF298, dont le génome fait 3,0 Mpb et permet la croissance en milieu minimal, pour générer des mutants d'insertion ou de délétion et les avons suivis dans plusieurs conditions distinctes afin de déterminer si cette souche possédait encore des gènes facultatifs. Ce mémoire traite

principalement des résultats obtenus dans un milieu riche et défini. Les données ainsi générées suggèrent que le génome de cette souche peut être réduit considérablement, puisque plus de 73 % de ses séquences codantes ne semblent pas être importantes en milieu riche. Une partie de ces éléments facultatifs peut cependant être impliquée pour la survie dans certaines conditions, telles qu'en présence d'une source de carbone autre que le glucose ou en absence d'oxygène. Lors de la réduction subséquente, il faudra donc déterminer le niveau d'adaptabilité désiré pour la souche.

Mots-clés : biologie synthétique, châssis cellulaire minimal, réduction génomique, gènes facultatifs, *Escherichia coli*

REMERCIEMENTS

J'aimerais d'abord remercier mon directeur de recherche Sébastien Rodrigue, puisque sans lui, je n'aurais pas eu la chance d'explorer la discipline de la biologie synthétique et puisqu'il m'a si bien supporté dans mes projets. J'aimerais aussi remercier mon codirecteur Pierre-Étienne Jacques ainsi que mes conseillers Vincent Burrus et Pascale Beauregard pour leurs conseils tout au long de ma maîtrise et pour l'évaluation de mon mémoire. Je remercie également tous les membres de mon laboratoire pour leur soutien technique et moral. Je remercie aussi mes parents pour toute l'aide qu'ils m'ont apportée, autant sur le plan personnel que financier. Finalement, je tiens à remercier Suzie Lachance, ma conjointe, ainsi que Sky Grenier pour leur constant appui et pour avoir égayé ma vie durant les sept dernières années.

TABLE DES MATIÈRES

SOMMAIRE.....	iv
REMERCIEMENTS	vi
TABLE DES MATIÈRES.....	vii
LISTE DES ABRÉVIATIONS	x
LISTE DES TABLEAUX	xii
LISTES DES FIGURES	xiii
CHAPITRE 1.....	1
1. INTRODUCTION.....	1
1.1 Biologie synthétique.....	1
1.2 Essentialité.....	3
1.2.1 Définition d’essentiel.....	3
1.2.2 Cas spéciaux d’essentialité	4
1.2.2.1 Létalité synthétique.....	4
1.2.2.2 Essentialité conditionnelle.....	4
1.2.2.3 Autres cas	5
1.3 E. coli.....	5
1.3.1 Description d’E. coli.....	5
1.3.2 Le génome essentiel d’E. coli.....	6
1.3.3 La réduction du génome d’E. coli	6
1.4 Mutagénèses	8
1.4.1 Mutagénèses dirigées.....	8
1.4.2 Mutagénèses aléatoires	9
1.4.2.1 Mutagénèse par insertions	9
1.4.2.2 Mutagénèse par délétions	11
1.5 Contexte, hypothèse et objectif de recherche	13
CHAPITRE 2.....	15
2. MATÉRIEL ET MÉTHODES	15
2.1 Souches et conditions de culture	15

2.2 Construction des plasmides	15
2.3 Conjugaison	17
2.4 Mutagenèse par insertions	18
2.5 Mutagenèse par délétions	18
2.6 Compte d'unités formant des colonies	19
2.7 PCR.....	19
2.8 Préparation des banques de séquençage	20
2.8.1 Séquençage des mutants d'insertion.....	21
2.8.2 Séquençage des mutants de délétion	21
2.9 Analyses bio-informatiques.....	22
2.9.1 Traitement des données de mutagenèse par insertions.....	22
2.9.2 Traitement des données de mutagenèse par délétions	23
Chapitre 3	25
3. RÉSULTATS	25
3.1 Mutagenèse par insertions	25
3.1.1 Validations de la méthode	25
3.1.2 Génération des mutants d'insertion	28
3.1.3 Séquençage des mutants d'insertion.....	28
3.1.4 Analyses des mutants d'insertion	30
3.1.4.1 Saturation de la mutagenèse	30
3.1.4.2 Identification des gènes importants	36
3.1.4.3 Identification des gènes importants avec le TraDIS toolkit	38
3.2 Mutagenèse par délétions	42
3.2.1 Génération des mutants de délétion.....	42
3.2.2 Confirmation des mutants de délétion	43
3.2.3 Séquençage des mutants de délétion	46
3.2.4 Analyses des mutants de délétion.....	47
3.2.4.1 Saturation de la mutagenèse	48
3.2.4.2 Sommaire de la mutagenèse	55
Chapitre 4	57
4. DISCUSSION.....	57

4.1 Mutagénèse par insertions	58
4.2 Mutagénèse par délétions	60
Chapitre 5	62
5. CONCLUSION	62
ANNEXE 1	64
Détails sur la construction des plasmides	64
pFG036	65
pFG051	66
pFG058	67
pFG057	68
ANNEXE 2	69
Gènes classés importants présents dans les délétions identifiées pour les jours 2 à 7.	69
ANNEXE 3	71
Analyse comparative d'essentialité en milieu riche ou minimal avec suivi en turbidostat ou par passages.	71
BIBLIOGRAPHIE	74

LISTE DES ABRÉVIATIONS

A	Adénine
ADN	Acide désoxyribonucléique
ARN	Acide ribonucléique
ARNg	ARN guide
ATP	Adénosine triphosphate
C	Cytosine
C _t	<i>Threshold cycle</i>
CDS	<i>Coding DNA Sequence</i>
COG	<i>Cluster of Orthologous Groups of proteins</i>
DAP	Acide diaminopimélique
DO ₆₀₀	Densité optique à 600 nm
dATP	Désoxyadénosine triphosphate
dNTP	Désoxyribonucléotide triphosphate
ES	<i>End sequence</i>
g	Gramme
G	Guanine
h	Heure
L	Litre
LB	<i>Lysogeny Broth</i>
ME	<i>Mosaic end</i>
MOPS	3-(N-morpholino)propanesulfonic acid
N	A, T, G ou C
nm	Nanomètre
pb	Paire de bases
PCR	Polymerase chain reaction

qPCR	PCR quantitative
R	A ou G
s	Seconde
T	Thymine
UFC	Unité formant des colonies
W	A ou T
Y	C ou T
2-2-BP	2,2'-Bipyridine

LISTE DES TABLEAUX

Tableau 1. Organismes au génome réduit	2
Tableau 2. Souches d'E. coli au génome réduit	7
Tableau 3. Plasmides utilisés dans cette étude	16
Tableau 4. Bruits lors du séquençage	37
Tableau 5. Filtre de couverture minimale des sites d'insertion	38
Tableau 6. Résumé de l'analyse TraDIS	39

LISTES DES FIGURES

Figure 1. Spectre d'importance d'un gène.....	3
Figure 2. Insertion d'un transposon.....	10
Figure 3. Mutagenèse par transposons avec pBAM1.....	11
Figure 4. Délétion par double transposition.	12
Figure 5. Confirmation de la sélection des mutants par compte d'UFC	26
Figure 6. Confirmation de la sélection des mutants par PCR.....	27
Figure 7. Préparation de la banque de séquençage des mutants d'insertion.....	29
Figure 8. Motif d'insertion de la transposase Tn5.....	31
Figure 9. Taux d'insertion selon le contenu GC.....	31
Figure 10. Saturation du séquençage.....	32
Figure 11. Nouveaux sites détectés au cours du suivi	33
Figure 12. Motifs des sites détectés au cours du suivi.	34
Figure 13. Localisation des sites d'insertion dans le génome de la souche DGF298.....	35
Figure 14. Exemple des résultats de la mutagenèse par insertions sur le UCSC genome browser	36
Figure 15. Taux de croissance relatif des mutants de gènes classés facultatifs selon TraDIS ..	40
Figure 16. Fonctions des séquences codantes facultatives à DGF298	41
Figure 17. Mutagenèse par délétions aléatoires.....	43
Figure 18. Confirmation de la sélection des mutants par compte d'UFC	44

Figure 19. Confirmation des mutants de délétion par PCR.....	45
Figure 20. Préparation de la banque de séquençage des mutants de délétion	47
Figure 21. Abondance relative des délétions affectant le locus ompR.....	49
Figure 22. Saturation du séquençage de la mutagenèse par délétions.....	50
Figure 23. Artéfacts générés durant la préparation de la librairie de séquençage	51
Figure 24. Nouvelles délétions détectées au cours du suivi	52
Figure 25. Alignement des lectures	53
Figure 26. Localisation des délétions dans le génome de la souche DGF298.....	55
Figure 27. Longueur totale des délétions	56

CHAPITRE 1

1. INTRODUCTION

1.1 Biologie synthétique

La biologie synthétique est une discipline récente qui consiste principalement à programmer des systèmes biologiques pour accomplir des fonctions spécifiques parfois absentes dans la nature (Cambray *et al.*, 2011). Ainsi, différentes modifications du contenu génique peuvent être faites, comme l'ajout ou le retrait de certains gènes, afin de concevoir de nouvelles souches étant capables, par exemple, de produire de nouveaux médicaments ou des molécules d'intérêts comme des biocarburants. Ce domaine peut donc jouer un rôle dans différents secteurs, autant du point de vue de la recherche fondamentale qu'appliquée. Pour aider le développement de la biologie synthétique, il est essentiel d'obtenir davantage de connaissances sur les règles permettant la conception et la modification extensive de génomes. Ceci apportera une nouvelle dimension aux récentes avancées dans la synthèse chimique complète de génomes et dans les modifications génétiques à haut débit (Wang *et al.*, 2009 ; Wang *et al.*, 2012 ; Hutchison III *et al.*, 2016 ; Ostrov *et al.*, 2016). Il est donc important de jumeler ce domaine à celui de la biologie des systèmes, qui consiste à utiliser différentes méthodes « omiques » et ainsi générer de nouvelles données qui permettent, une fois combinées, d'avoir une compréhension plus complète de l'organisme d'intérêt, ce qui permet ensuite de mieux le programmer.

Un des axes de la biologie synthétique vise à concevoir un châssis cellulaire minimal, qui est une cellule contenant uniquement le matériel génétique essentiel, ce qui réduit sa complexité et la difficulté de sa reprogrammation. Plusieurs projets visant l'obtention d'une telle cellule par réduction génomique sont en cours à travers le monde, et ce avec plusieurs organismes différents (tableau 1 ; Martínez-García and Lorenzo, 2016). Les organismes choisis présentent certains

avantages comme des propriétés industrielles intéressantes, un génome ayant déjà une faible taille ou une facilité à modifier son bagage génétique.

Tableau 1. Organismes au génome réduit

Organisme	Taille du génome initial	% retiré	Référence
<i>Escherichia coli</i>	4,6 Mpb	39 %	(Iwadate <i>et al.</i> , 2011)
<i>Mycoplasma mycoides</i>	1,2 Mpb	58 %	(Hutchison III <i>et al.</i> , 2016)
<i>Bacillus subtilis</i>	4,2 Mpb	36 %	(Reuß <i>et al.</i> , 2017)

Le plus petit génome permettant la vie cellulaire autonome obtenu à ce jour est celui de JCVI-syn3.0, un dérivé de *Mycoplasma mycoides* doté d'un seul chromosome circulaire de 531 Kpb portant 473 gènes (Hutchison III *et al.*, 2016). Bien que l'obtention de ce premier génome minimal soit une étape importante, d'autres doivent être obtenus à partir d'organismes variés, afin de bien comprendre les règles régissant l'architecture d'un organisme vivant. En effet, le contenu du génome obtenu suite au processus de réduction est dépendant de l'organisme sélectionné initialement, de l'ordre chronologique des délétions et du milieu de culture utilisé. Un autre projet de réduction à partir de *M. mycoides* pourrait alors mener à un produit différent de JCVI-syn3.0, selon le choix et l'ordre chronologique des délétions, ainsi que les conditions de croissance sélectionnées. Un organisme simplifié présentera donc des propriétés qui le distingueront des autres organismes au génome minimal. Il serait, par exemple, étonnant que le produit de réduction d'*Escherichia coli*, une bactérie Gram négatif ayant donc deux membranes et une paroi, soit le même que celui de *M. mycoides*, un mycoplasme possédant seulement une membrane.

1.2 Essentialité

1.2.1 Définition d'essentiel

Le concept de réduction génomique sous-entend qu'une partie du génome de l'organisme choisi n'est pas essentiel à sa survie dans les conditions utilisées. Bien qu'à priori triviale, la définition de ce que l'on qualifie d'essentiel n'est pas toujours la même à travers la communauté scientifique. En effet, certains classent les gènes comme étant strictement essentiels ou facultatifs, alors qu'il pourrait être plus juste d'employer un certain spectre d'importance (figure 1 ; Hutchison III *et al.*, 2016). À l'une des extrémités, il y aurait des gènes qui sont absolument nécessaires à la cellule et qui ne peuvent en aucun cas être retirés sans causer la mort de cette dernière. Si nous nous éloignons légèrement de cet extrême, nous pourrions retrouver des gènes dont le retrait peut être viable, mais dont le taux de croissance associé serait si faible que le mutant de délétion correspondant pourrait ne pas sembler être obtainable expérimentalement et ne serait de toute façon pas utile en laboratoire. Ensuite, il y aurait des gènes dont la délétion affecte le taux de croissance, mais de manière tolérable selon le contexte expérimental. Pour ces derniers, il faut être très conservateur lors d'une réduction génomique, puisque le retrait de nombreux de ces gènes peut causer des effets cumulatifs (Hutchison III *et al.*, 2016). Puis, il y a les gènes facultatifs dont la délétion n'affecterait pas de manière significative la croissance de la cellule. Finalement, nous pourrions même retrouver des gènes pouvant, selon l'environnement des cellules, diminuer la vitesse de croissance, comme un gène codant pour une protéine qui importe des composés à effets délétères. Le retrait de ces gènes serait donc bénéfique dans ce type de conditions.

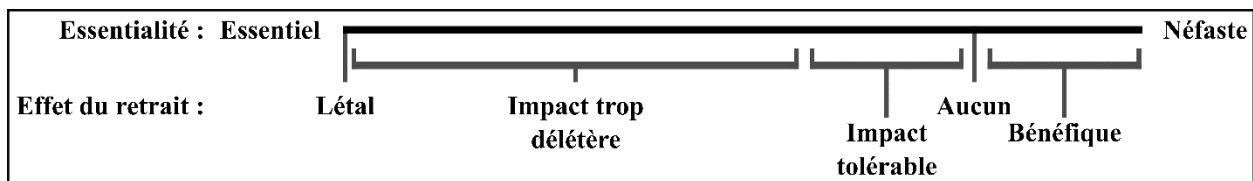


Figure 1. Spectre d'importance d'un gène. L'importance d'un gène se situe dans un spectre allant d'essentiel à néfaste.

1.2.2 Cas spéciaux d'essentialité

1.2.2.1 Létalité synthétique

La létalité synthétique implique de la redondance pour une fonction essentielle. Certains gènes liés à une fonction essentielle peuvent être inactivés, mais pas de manière simultanée, puisque l'activité en question doit être assurée dans les cellules. Les loci ribosomiaux en sont un exemple car leur nombre peut être changé sans trop affecter la cellule, mais ils ne peuvent être tous retirés puisque la cellule a besoin de ribosomes pour survivre dans toutes conditions (Gyorffy *et al.*, 2015). Un intérêt particulier doit être accordé à l'étude de ces cas lors de la simplification d'un génome pour bien choisir les gènes qui seront enlevés ou conservés et ainsi maximiser la réduction. Ce type d'essentialité peut être investigué avec une mutagénèse combinatoire, où l'on retire ou interrompt plusieurs gènes à la fois (Butland *et al.*, 2008).

1.2.2.2 Essentialité conditionnelle

Alors que de nombreux gènes seront toujours importants pour la cellule, plusieurs le sont seulement dans certaines conditions spécifiques. Ces cas d'essentialité conditionnelle soulignent l'importance de tester plusieurs conditions lors d'un projet de réduction génomique afin d'identifier les gènes importants pour créer une cellule réduite adaptée à nos besoins.

Un autre cas d'essentialité conditionnelle concerne plutôt certaines combinaisons de gènes, comme les systèmes toxine antitoxine (Hayes, 2003). En effet, alors que la paire de gènes peut souvent être retirée du génome, on ne peut pas déléter seulement l'antitoxine sans causer la mort de la cellule. L'antitoxine pourrait donc être classée essentielle par erreur. Comme pour la létalité synthétique, une mutagénèse combinatoire peut être utilisée pour détecter ces combinaisons.

1.2.2.3 Autres cas

Outre les cas mentionnés précédemment, d'autres, reliés aux méthodes utilisées pour identifier les gènes essentiels, sont à considérer. Par exemple, si un gène a une très courte taille, il est plus probable qu'il soit, lors d'une mutagénèse par transposons, exempt d'insertion simplement par chance et soit alors classé essentiel (Gerdes *et al.*, 2003). De façon similaire, lors d'une mutagénèse par délétion précise à l'aide de recombinaison double, il est possible qu'on ne puisse pas obtenir un certain mutant à cause d'un problème technique, comme la présence d'une structure secondaire empêchant la recombinaison de la cassette d'ADN, et non à cause de l'essentialité du locus ciblé (Baba *et al.*, 2006). Finalement, il est aussi possible qu'un gène soit facultatif, mais paraisse important dû à des effets polaires, causés par la modification génomique, sur les gènes adjacents.

1.3 *E. coli*

1.3.1 Description d'*E. coli*

Isolée initialement de fèces d'individus en santé, *E. coli* est une bactérie Gram négatif (Escherich, 1988). Souvent commensale ou mutualiste, elle peut devenir pathogène en acquérant des facteurs de virulence par transferts horizontaux (Leimbach *et al.*, 2013). Elle possède une grande capacité d'adaptation grâce aux nombreux gènes acquis au cours de son évolution dans un habitat en constant changement. Elle peut donc croître dans des milieux très variés, pauvres ou riches, et se divise rapidement. Depuis sa découverte en 1885, cette bactérie est devenue l'organisme modèle principal en biologie moléculaire (Blount, 2015). Également, grâce à son utilisation extensive, de nombreux outils moléculaires ont été développés chez celle-ci, ce qui en fait un organisme très utile pour étudier différents processus cellulaires.

1.3.2 Le génome essentiel d'*E. coli*

Plusieurs études ayant pour but d'identifier l'ensemble des gènes essentiels chez *E. coli* ont été effectuées par mutagenèse ciblée ou aléatoire (Gerdes *et al.*, 2003 ; Baba *et al.*, 2006 ; Shan *et al.*, 2015).

En effet, Baba *et al.* (2006) ont tenté de retirer individuellement 4288 des 4390 séquences codantes de la souche BW25113 connues lors de l'écriture de l'article. Ceci pour déterminer si les souches résultantes pouvaient survivre dans un milieu riche non défini, le *Lysogenic Broth* (LB). Ils ont ainsi été capables d'obtenir 3833 mutants de délétion simple (ensemble nommé collection Keio), ce qui correspond à 88,9 % des séquences codantes connues aujourd'hui dans cette souche (Yamamoto *et al.*, 2009 ; Grenier *et al.*, 2014). Sans considérer les cas de létalité synthétique, ces résultats suggèrent que la taille du génome essentiel d'*E. coli* pourrait être comparable à celle de JCVI-syn3.0, soit moins de 0,6 Mpb contenant 480 séquences codantes. D'autres études mettant à profit des mutagenèses aléatoires par insertion de transposon semblent également indiquer que le génome essentiel d'*E. coli* contiendrait peu de séquences codantes, soit entre 620 et 800 selon Gerdes *et al.* (2003) ou Shan *et al.* (2015), respectivement.

1.3.3 La réduction du génome d'*E. coli*

Plusieurs laboratoires dans le monde ont débuté, de manière indépendante, la réduction du génome d'*E. coli* (Iwadate *et al.*, 2011 ; Hirokawa *et al.*, 2013 ; Karcagi *et al.*, 2016). Un tel intérêt est porté à cette tâche puisqu'en plus de l'utilité liée à la production d'un nouveau type de bactérie minimale, beaucoup de nouvelles connaissances sur les protéobactéries, et tous autres organismes, seront découvertes par l'entremise de ce projet. En effet, la réduction d'un génome pourrait par

exemple nous permettre de découvrir de nouveaux gènes ou fonctions qui sont essentiels aux cellules (Hutchison III *et al.*, 2016).

Cette quête est approchée de manière différente par chacune des équipes de recherche. En effet, ces différents projets varient dans les propriétés que doit conserver ou obtenir la souche au génome réduit, dans la manière d'identifier les régions à retirer et dans la technique utilisée pour y arriver. Ceci mène donc à la création de plusieurs souches au génome réduit dont le contenu génique, le taux de réduction et le phénotype varient considérablement (tableau 2). DGF298 est la souche qui a subi la plus grande réduction, tout en conservant sa vitesse de division initiale et la capacité de croître en milieu minimal (Hirokawa *et al.*, 2013). Ayant eu de la difficulté à réduire son génome à moins de 3,0 Mpb sans affecter la croissance de la souche, ses créateurs semblent croire que cette souche est très près d'être une cellule minimale d'*E. coli*. La différence entre cette dernière taille et les tailles hypothétiques des génomes essentiels telles que mentionnées précédemment pourrait être expliquée par la présence de nombreux gènes non essentiels, mais qui seraient importants pour avoir un taux de croissance compétitif. Une hypothèse complémentaire, mentionnée par Martínez-García et Lorenzo (2016), serait que le génome de la cellule soit limité à une certaine taille, sans tenir compte de son contenu.

Tableau 2. Souches d'*E. coli* au génome réduit

Nom	Taille du génome	Croissance en milieu minimal	Vitesse de croissance diminuée	Référence
DGF298	3,0 Mpb	Oui	Non	(Hirokawa <i>et al.</i> , 2013)
LD33 α	2,8 Mpb	Non testé	Oui	(Iwadate <i>et al.</i> , 2011)
MDS69	3,7 Mpb	Oui	Oui	(Karcagi <i>et al.</i> , 2016)

1.4 Mutagénèses

De nombreux types de mutagénèses existent avec des objectifs variés, comme la création de génomes avec un code alternatif, l'optimisation de souches pour une certaine fonction industrielle ou l'identification des gènes facultatifs dans une condition donnée (Baba *et al.*, 2006 ; Julleson *et al.*, 2015 ; Shan *et al.*, 2015 ; Ostrov *et al.*, 2016). Plusieurs de ces types de mutagénèses peuvent être utilisés pour identifier le contenu facultatif du génome d'un organisme. Le principe est d'inactiver un gène et d'observer le phénotype résultant. Ceci peut être fait à l'aide de méthodes dirigées qui ont l'avantage de cibler précisément un gène, mais qui sont souvent limitées en termes de nombre de gènes pouvant raisonnablement être ciblés dans une expérience. Il existe aussi des méthodes aléatoires qui présentent l'avantage de perdre le biais inhérent aux méthodes dirigées et qui permettent de sonder un grand nombre de gènes simultanément. Ces dernières méthodes doivent tout de même avoir une efficacité élevée pour couvrir l'ensemble du génome.

1.4.1 Mutagénèses dirigées

Plusieurs méthodes permettent de retirer précisément, avec ou sans cicatrice, une partie du génome (Pósfai *et al.*, 1999 ; Datsenko and Wanner, 2000 ; Herring *et al.*, 2003). Elles impliquent généralement un mécanisme de recombinaison, en utilisant la machinerie de l'hôte ou un système exogène comme celui du phage lambda (Datsenko and Wanner, 2000). C'est ainsi que Baba *et al.* (2006) ont créé la collection Keio comprenant 3833 mutants de délétion simple. En effet, ils ont mis à profit les protéines Beta, Gam et Exo du phage lambda et ont tenté de retirer la plupart des gènes d'*E. coli*, de manière individuelle, en les remplaçant par une cassette comprenant un gène de résistance à un antibiotique, qui a permis de sélectionner les mutants obtenus.

1.4.2 Mutagénèses aléatoires

1.4.2.1 Mutagénèse par insertions

Une technique maintenant largement utilisée qui permet d'effectuer une mutagénèse par insertions aléatoires repose sur l'utilisation d'éléments génétiques mobiles, les transposons de type « couper-coller » (Reznikoff, 2008). Ceux-ci peuvent être considérés comme des parasites génomiques qui ont été transformés en outils moléculaires très utiles. Normalement, ces transposons sont minimalement constitués d'un gène codant pour une transposase et de deux régions identiques, mais inverses, qui le bordent (figure 2). On y retrouve aussi des gènes de résistances antibiotiques et d'autres impliqués dans la régulation de la transposition. Quand la transposase est produite, deux unités reconnaissent chacun des sites aux extrémités du transposon, dimérisent et coupent de façon franche de chaque côté. Il y a donc d'une part le génome qui doit être réparé par l'hôte et de l'autre, un transposome, formé d'au moins deux unités de transposase et du transposon, qui cherche alors une cible d'ADN pour y insérer le transposon. Selon celui-ci, la cible, dupliquée lors de l'insertion, varie en longueur et peut être très stricte au niveau de sa composition, comme pour les transposons mariner (mesurant 2 pb, et constituée des nucléotides TA), ou plutôt dégénérée, comme dans le cas du transposon Tn5 (mesurant 9 pb, et constituée préférentiellement des nucléotides GNTYWRANC) (figure 2 ; Robertson and Lampe, 1995 ; Goryshin *et al.*, 1998).

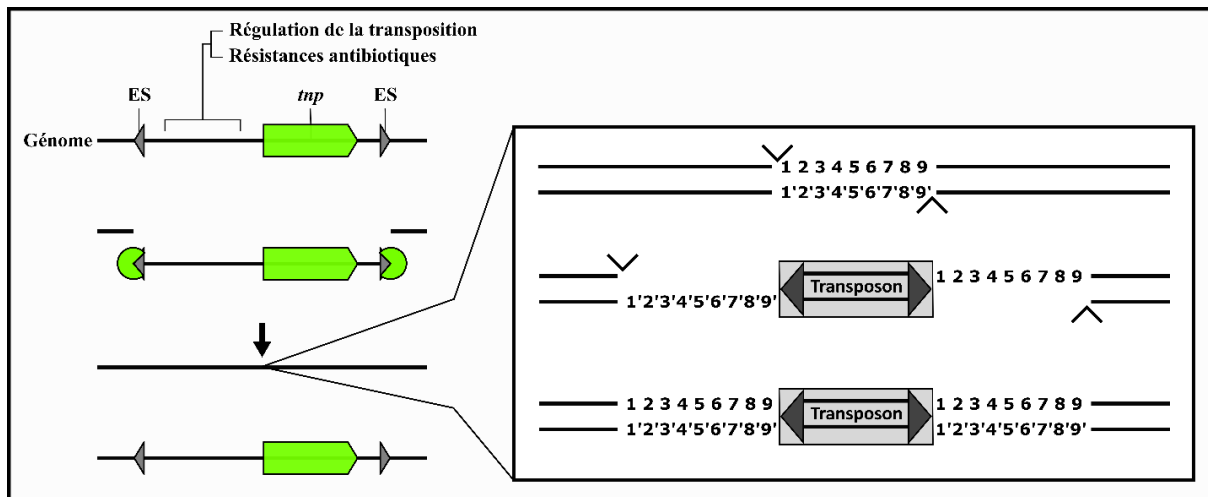


Figure 2. Insertion d'un transposon. Dans la nature, un transposon est formé d'au moins un gène codant pour une transposase (*tnp*), et de quelques autres gènes impliqués dans la sélection ou la régulation du transposon, le tout bordé de sites (ES). Lors d'un événement de transposition, au moins deux unités de transposase vont lier chaque site ES, dimériser et cliver de chaque côté. Ensuite le transposon est inséré à un autre endroit en causant une duplication du site d'insertion de longueur variable selon la transposase utilisée. Dans le cas du transposon Tn5, 9 pb sont dupliqués.

De nombreux outils moléculaires ont été construits en mettant à profit les transposons coupe-coller (Chiang and Rubin, 2002 ; Martínez-García *et al.*, 2011). Par exemple, la méthode EZ-tn5 (Lucigen) consiste à d'abord assembler des transposomes *in vitro* à l'aide d'un gène de résistance bordé de sites que reconnaît la transposase Tn5 et cette même transposase purifiée. Les transposomes formés peuvent ensuite être livrés par électroporation dans des cellules pour que le transposon puisse être inséré dans le génome et ainsi créer des mutants. Ceux-ci peuvent être soumis à une sélection dans des conditions bien définies, puis séquencés pour identifier les mutants d'insertions ayant survécu. Une autre méthode consiste à utiliser un plasmide portant un transposon et codant pour la transposase, tel que pBAM1 (figure 3 ; Martínez-García *et al.*, 2011). Ce plasmide doit être livré dans la cellule, préférentiellement par conjugaison pour obtenir davantage de mutants. Une fois présent, la transposase est produite et formera un transpososome en excisant le transposon du plasmide. Après quoi le transposon est inséré dans le génome et les mutants sont sélectionnés et séquencés. Le plasmide utilisé est rapidement éliminé, car il ne peut être répliqué dans la cellule réceptrice.

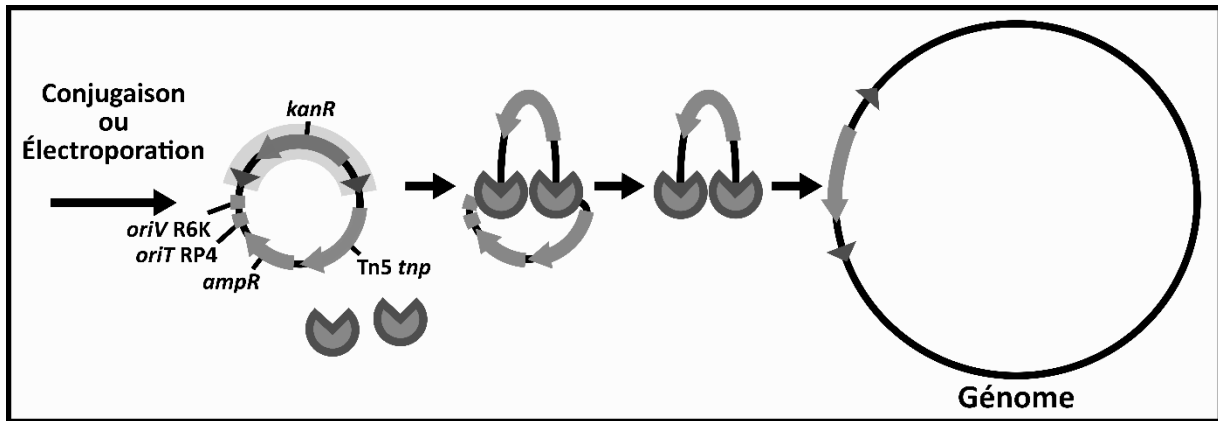


Figure 3. Mutagenèse par transposons avec pBAM1. La mutagenèse par transposons avec un plasmide suicide comme pBAM1 se sépare en deux principales étapes, la livraison du plasmide et la transposition (Martínez-García *et al.*, 2011). D’abord, à l’aide de la conjugaison, ou toutes autres méthodes de transformation, le plasmide est introduit dans la cellule d’intérêt. Puisque le plasmide a une origine de répllication dépendante de la protéine pi et que celle-ci n’est pas présente dans la cellule, il sera éventuellement perdu lors de l’ultérieure mise en culture des cellules. Avant de le perdre, chez de nombreuses cellules, une transposase, produite à partir du plasmide, formera un transposome avec un transposon synthétique aussi présent dans le plasmide et insérera ce transposon aléatoirement dans le génome.

1.4.2.2 Mutagenèse par délétions

Plusieurs méthodes ont été développées pour générer des délétions aléatoires (Yu *et al.*, 2002 ; Goryshin *et al.*, 2003). Une de ces méthodes consiste à d’abord créer des mutants d’insertion avec deux transposons distincts contenant chacun un site de recombinaison *loxP* et d’ensuite produire une délétion avec la recombinaise CRE et les deux sites. L’autre méthode implique plutôt une double transposition avec un transposon plus complexe. De façon semblable aux méthodes d’insertions aléatoires vues précédemment, le transposon est d’abord inséré dans le génome de la cellule. Par contre, ce transposon code pour une seconde transposase qui, lorsque son expression est induite, reconnaît deux sites spécifiques situés dans le transposon et forme un second transposome (figure 4A). À cause de l’orientation des sites reconnus, le génome entier

est alors utilisé comme transposon. Le seul endroit où il peut être inséré est dans sa propre séquence, ce qui peut causer une inversion ou la fission du chromosome, selon l'orientation des brins d'ADN. Dans le deuxième scénario, une délétion est obtenue puisqu'une des deux structures n'a pas d'origine de réplication et sera donc perdue.

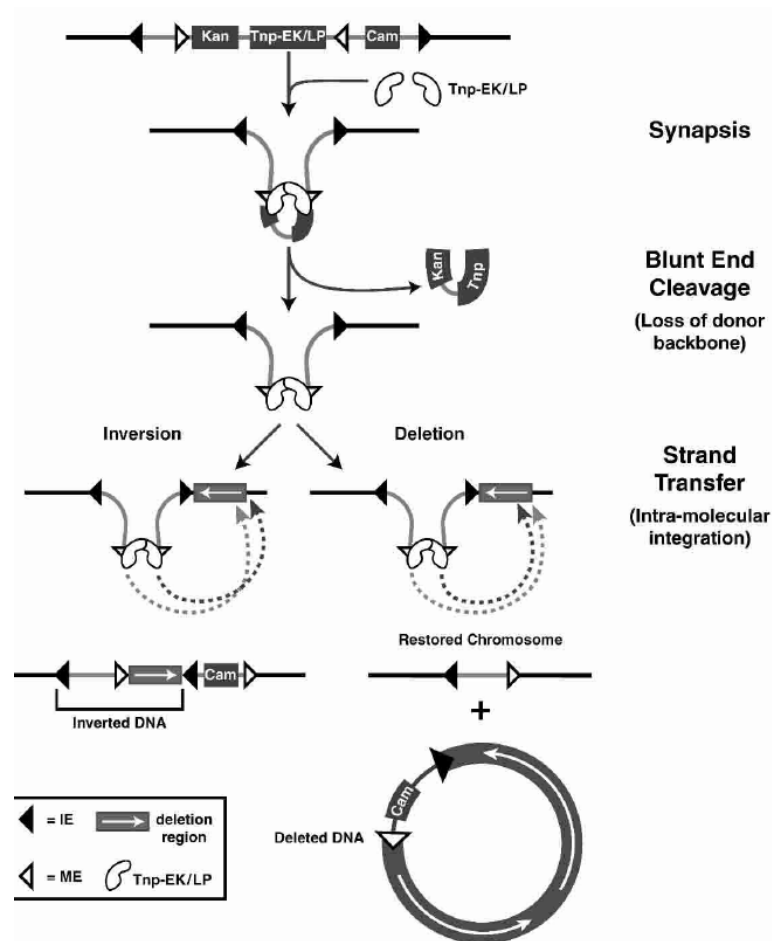


Figure 4. Délétion par double transposition. Le transposon, une fois inséré dans le génome, permet, lorsqu'induit, de produire une seconde transposase, reconnaissant des ES différents, qui forme de nouveau un transposome, mais cette fois-ci avec le génome entier comme transposon. Plusieurs conclusions peuvent être obtenues suite à ce processus. En effet, une délétion ou une inversion peuvent se produire d'un côté ou de l'autre du transposon, en fonction du site d'insertion du second événement de transposition. Seules les cellules ayant fait une délétion du côté droit du transposon sont sélectionnées en conservant seulement des mutants qui ne possèdent plus de marqueur de résistance. Modifié de (Goryshin *et al.*, 2003).

Jusqu'à présent, cette méthode a été testée chez *E. coli* sans être utilisée à son plein potentiel. Ainsi, seulement 42 délétions distinctes mesurant de quelques centaines de paires de bases à plus de 100 Kpb ont été générées (Goryshin *et al.*, 2003).

1.5 Contexte, hypothèse et objectif de recherche

Les avancées dans la synthèse chimique de génomes bactériens complets et dans les modifications génétiques à haut débit des dernières années témoignent de l'effervescence de la biologie synthétique. Les récents progrès rendent possible la création de tout nouveaux organismes dotés de nouvelles capacités utiles à l'humanité. Malgré la présente capacité de synthèse d'ADN, le manque de connaissances sur les règles régissant la composition et l'organisation d'un génome limite notre efficacité à créer un génome artificiel de novo. Dans le laboratoire du professeur Sébastien Rodrigue, nous étudions donc la biologie des systèmes permettant d'acquérir un tel savoir destiné à être appliqué avec une perspective de biologie synthétique. L'un des axes du laboratoire est ainsi orienté vers les études « omiques » d'organismes existants, dont *Mesoplasma florum* et *E. coli*, avec comme objectifs parallèles de réduire leur contenu génomique au minimum, de les comparer aux autres génomes minimaux obtenus naturellement ou de façon synthétique et d'ainsi dévoiler les règles de l'architecture génomique.

La présente étude est liée à la réduction du génome d'*E. coli* pour en faire un châssis cellulaire minimal dérivé d'une protéobactérie. Malgré que cet organisme ne possède pas un des plus petits génomes, il offre l'avantage d'être facile à manipuler génétiquement, facteur majeur dans une telle entreprise. Bien que certaines souches d'*E. coli* au génome réduit ont déjà été obtenues, elles sont loin de toutes partager les mêmes délétions, qui elles-mêmes incluent une faible portion des gènes non essentiels précédemment identifiés. Nous avons donc émis l'**hypothèse**

que le génome de la souche simplifiée DGF298, la plus réduite à ce jour (3,0 Mpb) qui a conservé un taux de croissance similaire à sa souche parente et qui peut croître en conditions minimales, porte encore des gènes facultatifs.

L'objectif principal du projet était d'identifier expérimentalement le contenu facultatif du génome de la souche DGF298 grâce à des méthodes de mutagenèse par insertions ou délétions que j'ai développées et optimisées.

La survie des mutants a été testée en milieu riche ou minimal, mais ce mémoire porte principalement sur les résultats obtenus en conditions riches.

CHAPITRE 2

2. MATÉRIEL ET MÉTHODES

2.1 Souches et conditions de culture

Les souches *E. coli* EC100Dpir+ (Epicentre) et DH5 α pir+ (généreux don du laboratoire de Vincent Burrus) ont servi aux clonages plasmidiques. *E. coli* MFDpir (Ferrières *et al.*, 2010), auxotrophe pour le DAP, a été utilisée comme cellule donneuse pour les conjugaisons et *E. coli* DGF298 (Hirokawa *et al.*, 2013) était la souche d'intérêt pour l'étude d'essentialité. Elles étaient toutes incubées à 30°C pour leur croissance. Le milieu *Lysogenic Broth* (LB) a été utilisé pour les clonages, alors que les milieux définis ont servi pour les mutagénèses. Le protocole suivi pour faire ces derniers milieux est disponible à l'adresse <https://www.genome.wisc.edu/resources/protocols/ezmedium.htm> ou <https://www.genome.wisc.edu/resources/protocols/mopsminimal.htm>. Pour l'entreposage à long terme, les cellules ont été conservées à -80°C dans du milieu LB contenant 25 % (v/v) glycérol. Les antibiotiques ou autres additifs ont été utilisés aux concentrations finales suivantes : chloramphénicol 34 μ g/mL, tétracycline 15 μ g/mL, spectinomycine 50 μ g/mL, kanamycine 50 μ g/mL, acide diaminopimélique (DAP) 57 μ g/mL et 2,2'-bipyridine (2-2-BP) 62 μ g/mL.

2.2 Construction des plasmides

Les plasmides utilisés dans cette étude sont répertoriés dans le tableau 3. Plus de détails sur leur conception peuvent être trouvés dans l'annexe 1. Brièvement, leur assemblage s'est fait en

utilisant le mélange Gibson NEBuilder HiFi DNA Assembly Master Mix (NEB) avec les différents fragments, obtenus par amplifications PCR ou digestions enzymatiques. Les digestions ont été effectuées avec les enzymes de restrictions provenant de la compagnie NEB. Les souches EC100Dpir+ ou DH5 α pir+ ont été utilisées pour le clonage des plasmides. La souche DH5 α pir+ était choisie lorsque le plasmide contenait un gène de résistance à la spectinomycine. Les clonages ont été faits par électroporation des cellules gardées sur glace et lavées deux fois à l'eau stérile déionisée, avec la totalité du produit d'assemblage Gibson purifié. Tous les plasmides ont été confirmés à l'aide de digestions par enzymes de restriction et/ou par séquençage de type Sanger de la plateforme de séquençage et de génotypage du centre de recherche du centre hospitalier de l'Université Laval. Les purifications requises ont été faites avec les billes magnétiques HighPrep PCR (MagBio Genomics), à l'exception des digestions par enzymes de restriction, qui ont été purifiées avec le DNA Clean & Concentrator (Zymo Research). Toutes les réactions enzymatiques ou étapes de purification ont été faites en suivant les recommandations des fabricants.

Tableau 3. Plasmides utilisés dans cette étude

Nom	Résistance	Fonction	Référence
pFG036	Tétracycline	Répression de la transposition dans la cellule donneuse	Cette étude
pFG051	Spectinomycine	Plasmide suicide de transposition	Cette étude
pFG058	Kanamycine	Plasmide suicide de délétion	Cette étude
pFG057	Tétracycline	Plasmide de contre-sélection Cas9	Cette étude

pSC189	Kanamycine Ampicilline	Matrice pour amplifier le gène <i>tnp</i> - mariner	(Chiang and Rubin, 2002)
pBAM1	Kanamycine Ampicilline	Matrice pour amplifier le gène <i>tnp</i> - Tn5, <i>oriV</i> -R6K, <i>oriT</i> - RP4	(Martínez-García <i>et al.</i> , 2011)
pCas9	Chloramphénicol	Matrice pour amplifier le gène <i>cas9</i> et l'origine de réplication du plasmide p15A	(Jiang <i>et al.</i> , 2013)
pUC19	Ampicilline	Matrice pour amplifier l'origine de réplication du plasmide colE1	(Norlander <i>et al.</i> , 1983)

2.3 Conjugaison

Les conjugaisons ont été effectuées avec la souche donneuse MFDpir contenant les plasmides à conjugués, ainsi que leur plasmide de soutien si nécessaire. DGF298 constituait la bactérie receveuse lors de ces expériences. En moyenne, un volume de 220 µL d'une culture de cellules receveuses à une densité optique à 600nm (DO₆₀₀) de 0,4 fut mélangé à 5 fois plus de cellules donneuses. Les quantités de cellules furent estimées avec le logiciel disponible à l'adresse <http://www.genomics.agilent.com/biocalculators/calcODBacterial.jsp> . Ensuite, deux lavages furent effectués avec le milieu de culture pour éliminer les antibiotiques présents lors de la croissance des cultures. Après quoi les cellules ont été resuspendues dans le milieu de culture et déposées sur une gélose du même milieu, mais contenant du DAP. Il s'agit alors d'une

gouttelette de 100 μ L avec une très haute densité cellulaire, soit environ 4 milliards de cellules par mL. Lorsque davantage d'exconjugants étaient désirés, plusieurs gouttelettes identiques ont été déposées parallèlement. La gélose portant les gouttelettes a ensuite été placée sous une hotte stérile pendant 1 heure pour que l'eau s'évapore et que les cellules forment des disques plats et compacts. Après quoi la gélose fut laissée 2 heures à température pièce pour laisser place à la conjugaison. Les cellules ont ensuite été resuspendues, lavées une fois avec le milieu de croissance pour retirer le DAP et mises en culture pour qu'un suivi au cours du temps soit effectué.

2.4 Mutagénèse par insertions

La mutagénèse par insertion de transposon a été effectuée en suivant le protocole de conjugaison avec la souche donneuse MFDpir portant les plasmides pFG036 et pFG051. 100 gouttelettes de conjugaison, décrites dans la section précédente, ont été utilisées. Celles-ci ont été mises en culture dans 1 L de milieu avec spectinomycine après la conjugaison. Le premier passage a été effectué en transférant 60 μ L de cette culture, ayant été incubée pendant 24 h, dans 10 mL de milieu frais avec spectinomycine. Les passages suivants ont été faits avec un inoculum de 10 μ L et le même temps d'incubation. Les échantillons pris au cours du suivi étaient préparés sous forme de culots bactériens sans surnageant faits avec 3 mL de culture et congelés à -80°C.

2.5 Mutagénèse par délétions

La première partie de la méthode consistait à faire une conjugaison avec une souche donneuse MFDpir portant les plasmides pFG036 et pFG058. Après la conjugaison, faite avec 10 gouttelettes, et le lavage, les cellules ont été mises en culture dans 100 mL de milieu avec kanamycine et sulfate de fer. Les passages ont été effectués en transférant 10 μ L de la culture

précédente, ayant été incubée pendant 24 h, dans 10 mL de milieu frais. Le chélateur 2,2'-bipyridine a été ajouté à la quatrième culture de mutants d'insertion, dont le milieu ne contenait alors pas de kanamycine, lorsque sa DO₆₀₀ a atteint 0,1. L'induction s'est ainsi poursuivie pendant 8 heures pour que les cellules puissent accomplir la deuxième transposition. Ensuite, elles ont été utilisées comme receveuses dans une seconde étape de conjugaison, impliquant 100 gouttelettes, avec la souche MFDpir contenant le plasmide pFG057. Les cellules furent alors mises en culture dans 1 L de milieu avec tétracycline pour sélectionner ce dernier plasmide et ainsi contre-sélectionner les cellules qui n'étaient pas des mutants de délétion. Les passages ont été effectués pendant 7 jours en transférant 10 µL de la culture précédente, ayant été incubée pendant 24 h, dans 10 mL de milieu frais.

2.6 Compte d'unités formant des colonies

Des comptes d'unités formant des colonies (UFC) furent utilisés pour estimer le nombre de chacun des types de cellules durant les suivis effectués suite aux mutagénèses. Pour ce faire, des dilutions sérielles ont été faites avec les échantillons. Ensuite, 5 µL de chacune des dilutions furent déposés sur des géloses contenant le milieu approprié et les antibiotiques permettant de sélectionner les cellules voulues, après quoi les géloses furent incubées à 30°C jusqu'à apparition de colonies. Celles-ci ont alors été comptées manuellement et les facteurs de dilution furent appliqués pour déterminer le nombre approximatif de chaque type de cellules.

2.7 PCR

Les réactions PCR ont été faites avec la polymérase VeraSeq 2.0 (Enzymatics), selon les recommandations du fabricant. Pour chaque PCR à partir d'une colonie bactérienne, un embout de pipette, ayant servi à piquer une colonie, fut mis quelques secondes dans le mélange à PCR

pour fournir la matrice d'ADN. Pour la visualisation des produits d'amplification sur gel électrophorèse, nous avons utilisé l'échelle 2-log ladder (NEB). Les PCR quantitatives ont été faites en ajoutant du colorant EvaGreen (Biotium) aux réactions et en utilisant le thermocycleur CFXconnect (Bio-Rad). Le nombre d'insertions par cellule a été calculé en utilisant la technique du $\Delta\Delta C_t$. Deux qPCR ont donc été faites en ciblant un gène contrôle, *rpoB*, et une partie du transposon. Ces réactions ont été faites avec chacun des mutants testés et un contrôle avec la partie du transposon ciblée d'intégrée dans son génome. Le calcul donnant le nombre d'insertions est le suivant :

$$\text{Nombre d'insertions} = 2^{((\Delta C_{t\text{mutant}}) - (\Delta C_{t\text{contrôle}}))} \text{ où } \Delta C_t = C_{t\text{transposon}} - C_{t\text{rpoB}}$$

2.8 Préparation des banques de séquençage

Les premières étapes de la préparation des banques de séquençage furent les mêmes pour les mutants d'insertion ou de délétion. D'abord, il a fallu décongeler sur glace chacun des échantillons, conservés sous forme de culots bactériens à -80°C. Ensuite, les cellules ont été resuspendues dans 300 µL de tampon de lyse du Quick-gDNA MiniPrep (Zymo Research), et soumises à une sonication avec le Bioruptor Plus (Diagenode), à force moyenne avec une alternance de 30 s en marche et 30 s en pause. Après quoi le reste du protocole fourni avec le Quick-gDNA MiniPrep (Zymo Research) a été suivi pour compléter la purification des fragments d'ADN. Ensuite, leurs extrémités ont été réparées avec le *End-Repair Mix HC* (Enzymatics), en utilisant tout ce qui fut récupéré, soit environ 10 µg d'ADN.

2.8.1 Séquençage des mutants d'insertion

Après avoir réparé les fins des fragments, les extrémités 3' de ceux-ci ont été adénylées avec la TaqB (Enzymatics) dans un mélange contenant seulement le tampon à PCR et du dATP, le tout incubé à 68°C pendant 30 minutes. Ensuite, une ligation avec la ligase T4 (Enzymatics) et l'adaptateur Nextera B d'Illumina, présent en excès (10 fois plus), a été effectuée. L'adaptateur a été conçu en hybridant deux oligonucléotides à une concentration de 40 µM dans un tampon d'hybridation (10 mM Tris NaCl pH 7,5, 50 mM NaCl). La réaction permettant l'hybridation consistait à chauffer le mélange à 98°C, puis à diminuer progressivement la température jusqu'à 4°C en 940 cycles. Suite à la ligation, une digestion a été effectuée avec les enzymes AvrII, XbaI et FseI (NEB) pour éliminer, si présents, les fragments issus du plasmide pFG051, après quoi une PCR permet d'obtenir les fragments prêts à être séquencés. La seule étape n'ayant pas été purifiée avec les billes magnétiques HighPrep PCR (MagBio Genomics) fut celle de digestion. Celle-ci fut plutôt purifiée avec le DNA Clean & Concentrator (Zymo Research). Avant d'envoyer les échantillons, nous avons confirmé la taille des fragments avec l'instrument Agilent Bioanalyzer 2100 et le Agilent High Sensitivity DNA Kit (Agilent) et la concentration avec le Synergy HT (BioTek) et le Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher). Les produits d'amplification ont ensuite été combinés, totalisant 75 µL à 2,3 ng/µL, et furent envoyés à la plateforme de séquençage du Centre d'innovation Génome Québec et Université de McGill pour être séquencés par l'appareil HiSeq 2000 (Illumina) en lectures simples de 50 pb.

2.8.2 Séquençage des mutants de délétion

Après avoir réparé les extrémités des fragments, une ligation intramoléculaire a été faite avec la ligase T4 Rapid (Enzymatics) et une concentration d'ADN de 1 ng/µL. Ensuite, deux PCR successives ont permis d'amplifier la portion du cercle correspondant au génome et d'y ajouter les adaptateurs requis au séquençage (figure 23). Avant d'envoyer les échantillons, nous avons

confirmé la taille des fragments avec l'instrument Agilent Bioanalyzer 2100 et le Agilent High Sensitivity DNA Kit (Agilent) et la concentration avec le Synergy HT (BioTek) et le Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher). Les produits d'amplification ont ensuite été combinés, totalisant 25 µL à 7,0 ng/µL, et furent envoyés à la plateforme de séquençage du CHUS pour être séquencés par l'appareil MiSeq (Illumina) en lectures en paire. 50 pb étaient allouées pour la première lecture et 100 pour la deuxième, celle-ci ne commençant pas directement dans le génome, mais plutôt dans la cicatrice laissée suite à la délétion.

2.9 Analyses bio-informatiques

2.9.1 Traitement des données de mutagenèse par insertions

L'extrémité des lectures brutes a d'abord été rognée selon leur qualité ou la présence d'adaptateur avec Trimmomatic, version 0.32 (Bolger *et al.*, 2014). La qualité des lectures a été inspectée avec l'outil FastQC, version 0.11.4 (Andrews, 2010). Les lectures ont ensuite été alignées sur le génome de DGF298 avec BWA, version 0.7.10 (Li, 2013). Celles avec une valeur d'alignement (MAPQ) au moins égale à 30 ont été conservées. Nous avons choisi la paire de bases centrale de la duplication de 9 pb comme position pour représenter chacun des sites. L'identification de la position des sites couverts par plus d'une lecture et le calcul des indices d'insertion, soit la couverture d'un gène normalisée par sa longueur, a été faite avec un script maison. Ceci en excluant les insertions présentent dans le début (<5 %) ou la fin (>85 %) des séquences codantes puisqu'elles pourraient ne pas affecter leur fonction. La couverture des sites a été normalisée en fonction du nombre total de lectures alignées sur le génome de DGF298 pour chacun des échantillons. La classification d'essentialité des gènes a ensuite été faite avec le script R *tradis_essentiality* disponible à l'adresse https://github.com/sanger-pathogens/Bio-Tradis/blob/master/bin/tradis_essentiality.R (Barquist *et al.*, 2016). Le motif d'insertion a été généré avec l'outil BLogo, disponible en ligne à <http://www.bioinformatics.org/blogo/cgi->

[bin/Blogo/Blogoform.pl](#), en utilisant 33 333 lectures, choisies aléatoirement, de chacun des 3 échantillons sélectionnés (Li *et al.*, 2008). La liste utilisée contenant les groupes fonctionnels pour chaque gène est disponible à l'adresse <http://www.microbesonline.org/cgi-bin/genomeInfo.cgi?Id=316407;export=tab>. Le taux de croissance relatif des gènes mutés a été calculé avec la formule suivante (van Opijnen *et al.*, 2009) :

$$W = \ln(N_{T3} \times D / N_{T1}) / \ln((1 - N_{T3}) \times D / (1 - N_{T1}))$$

où N_{T1} et N_{T3} représentent la fréquence des mutants du gène dans la population au premier et au dernier des temps sélectionnés, respectivement. D représente le facteur d'expansion entre les deux temps, lui-même calculé comme suit :

$$D = 2^{(\text{doublements effectués au temps 3} - \text{doublements effectués au temps 1})}$$

2.9.2 Traitement des données de mutagenèse par délétions

Les paires de lectures dont la deuxième portait l'extrémité attendue de la cicatrice laissée suite à la délétion ont été sélectionnées avec NovoBarcode, version 2.08 (Novocraft Technologies). Cette même région, précédant le site d'insertion, fut retirée des lectures. Ensuite, seules les paires dont la lecture 2 commençait par les nucléotides TA ont été conservées. Puis, leur extrémité a été rognée selon la qualité ou la présence d'adaptateur avec Trimmomatic, version 0.32 (Bolger *et al.*, 2014). La qualité des lectures a été inspectée avec l'outil FastQC, version 0.11.4 (Andrews, 2010). Les lectures ont ensuite été alignées sur le génome de DGF298 avec BWA, version 0.7.10 (Li, 2013). Celles avec une valeur d'alignement (MAPQ) au moins égale

à 30 ont été conservées. Nous avons choisi la paire de bases centrale de la duplication de 9 pb comme position pour les sites d'insertion Tn5 et la première paire de bases, sur le brin positif, pour les sites d'insertion mariner. Nous avons ensuite utilisé des scripts maison pour identifier la taille des inversions ou délétions, ces deux événements se distinguant par l'orientation des deux lectures.

Chapitre 3

3. RÉSULTATS

3.1 Mutagénèse par insertions

3.1.1 Validations de la méthode

Au début de nos travaux, nous nous demandions si le génome d'*E. coli* pouvait être réduit à moins de 3,0 Mpb, la taille du génome de la souche DGF298, tout en conservant un taux de croissance similaire à celui de la souche sauvage. Nous avons donc conçu le plasmide pFG051, semblable au plasmide pBAM1 (Martínez-García *et al.*, 2011), permettant de générer des mutants d'insertion pour déterminer si la souche portait encore des gènes facultatifs. Ce nouveau plasmide offrait la possibilité, contrairement à pBAM1, de pouvoir réprimer fortement l'expression du gène *tnp* grâce au plasmide pFG036, aussi conçu dans le cadre de cette étude. Également, pFG051 portait déjà un adaptateur de séquençage à l'une des extrémités de son transposon. Brièvement, la méthode utilisée consistait à livrer par conjugaison le plasmide suicide pFG051 de la souche donneuse MFDpir vers la souche receveuse DGF298. La transposase Tn5 était alors produite à partir de pFG051 et, après avoir excisé le transposon du même plasmide, allait l'insérer dans le génome de la cellule receveuse. Ce processus de transposition était inactif dans la cellule donneuse puisque celle-ci possédait également le plasmide pFG036 codant de façon constitutive pour un répresseur du gène *tnp*. Cette mutagénèse, une fois optimisée, permettait de transformer en moyenne 3 % des cellules receveuses en mutants d'insertion. En utilisant de grandes quantités de cellules, nous pouvons donc générer des millions de mutants.

Pour valider la présence de mutants et leur sélection spécifique durant un suivi, des comptes d'UFC ont été faits avec des échantillons pris à plusieurs intervalles. Les combinaisons d'antibiotiques choisis ont permis de suivre l'évolution de trois groupes de cellules : les donneuses, les mutants et les receveuses (avec ou sans transposon). La figure 5 montre un exemple d'un suivi des différents types cellulaires en turbidostat (Matteau *et al.*, 2015). Nous pouvons voir que les seules cellules semblant croître sont bien les receveuses ayant acquis un transposon puisque l'augmentation de la quantité de receveuses coïncide avec l'augmentation de la quantité de mutants. Nous pouvons aussi voir que les cellules donneuses sont graduellement éliminées avec les dilutions.

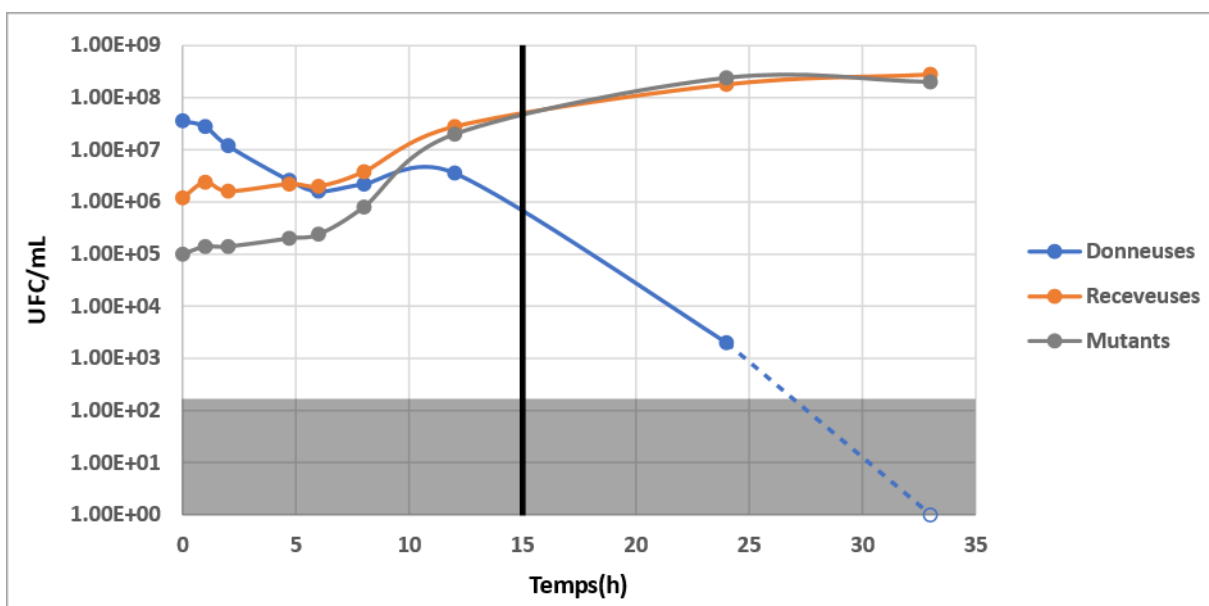


Figure 5. Confirmation de la sélection des mutants par compte d'UFC. Quantité de donneuses (DAP, tétracycline et spectinomycine ; bleu), de receveuses (orange) et de mutants (spectinomycine ; gris) en milieu riche au cours d'un suivi en turbidostat. La zone grise représente la limite de détection de la méthode de compte d'UFC et le début des dilutions est indiqué par la ligne noire.

Une confirmation par PCR sur colonie a également été faite avec 20 colonies isolées des temps 24 h ou 33 h, 10 provenant de géloses sans spectinomycine et 10 autres de géloses avec spectinomycine pour sélectionner les mutants. Le gène *aadA*, présent dans le transposon, fut la cible d'amplification. Tel que montré dans la figure 6, toutes les colonies testées correspondaient à des mutants.

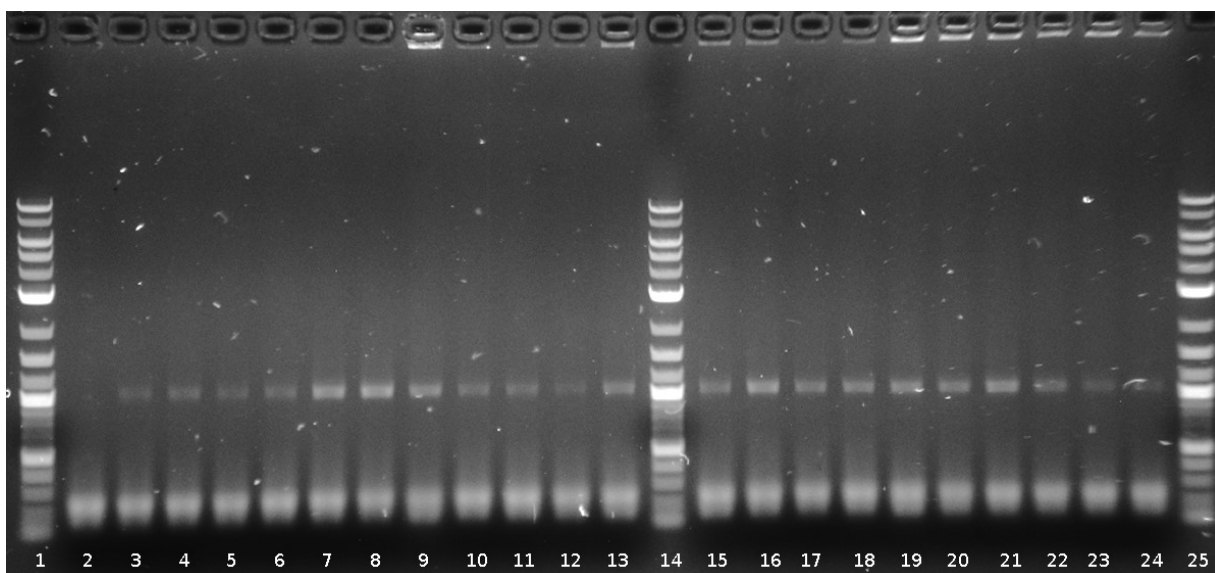


Figure 6. Confirmation de la sélection des mutants par PCR. Gel électrophorèse avec le produit d'amplification du gène *aad7* (1034 pb). Le puits 2 avait pour matrice de l'ADN génomique sans transposon et correspondait donc au contrôle négatif. Le puits 3 contenait le contrôle positif, avec le plasmide pFG051 comme matrice. Les puits 4 à 13 correspondaient aux colonies isolées de géloses avec spectinomycine et les puits 15 à 24, aux colonies isolées de géloses sans spectinomycine. Seul le contrôle négatif ne présentait pas de bande à 1034 pb.

Une mutagenèse par transposons vise la plupart du temps à générer des mutants avec seulement une insertion par cellule. Or, la méthode utilisée, due à son efficacité élevée, pourrait générer des cellules avec plusieurs transposons. Comme dernière vérification, nous avons donc utilisé la méthode de la PCR quantitative pour mesurer le nombre d'insertions dans 20 mutants isolés. De ceux-ci, 17 avait seulement un transposon et 3 en avait deux. Il semble donc que la majeure partie des mutants, soit environ 85 %, possédaient une seule insertion.

3.1.2 Génération des mutants d'insertion

Nous avons réussi à générer environ $6,0 \times 10^7$ mutants pour la mutagenèse faite pour cette étude. Cette dernière donnée fut estimée avec le compte d'UFC d'un échantillon pris au début du suivi. Cette méthode peut mener à une surestimation du nombre de mutants, puisque leur génération implique une incubation de deux heures. Par contre, la valeur obtenue ne devrait pas trop déroger de la réalité puisque les conditions utilisées pour les mutagenèses, soit à température pièce avec une forte densité cellulaire, sont peu favorables à la croissance.

En assumant que la transposase Tn5 insère le transposon de manière parfaitement aléatoire, le nombre de sites couverts peut être estimé avec une distribution de Poisson. Puisqu'un transposon peut être inséré au même endroit dans plusieurs cellules, le nombre attendu de sites couverts est inférieur au nombre de mutants. Nous estimions donc la quantité de sites couverts à $6,0 \times 10^6$, ce qui se traduit par une densité d'insertions estimée de $2,0 \times 10^3$ insertions/Kpb, les transposons pouvant s'insérer dans les deux orientations.

Suite à la mutagenèse initiale, les mutants furent sélectionnés avec de la spectinomycine en milieu défini riche et suivis pendant sept jours par passages journaliers.

3.1.3 Séquençage des mutants d'insertion

Du séquençage haut débit multiplexé a été utilisé pour identifier précisément l'emplacement du transposon dans les mutants échantillonnés au cours du suivi. Le transposon utilisé portait à l'une de ses extrémités l'adaptateur Nextera A (figure 7). Lors de la préparation de la banque de

séquençage, après avoir extrait, fragmenté, réparé et adénylé l'ADN, l'adaptateur Nextera B a été ajouté aux fragments (figure 7A-B). Ensuite, une PCR a permis d'amplifier majoritairement les séquences portant les deux adaptateurs distincts. En effet, les fragments contenant l'adaptateur A, présent à l'une des extrémités du transposon, étaient amplifiés de façon préférentielle puisque l'amorce A permettait d'amplifier un brin d'ADN portant la séquence pour l'hybridation de l'amorce B (figure 7C). Cela dit, il a fallu utiliser une température d'hybridation élevée pour diminuer le niveau d'amplification non spécifique. Nous avons choisi le séquençage de type Nextera puisqu'il a permis, en commençant le séquençage dans la partie génomique suivant l'adaptateur A, d'augmenter le nombre de paires de bases lues correspondantes aux sites d'insertions.

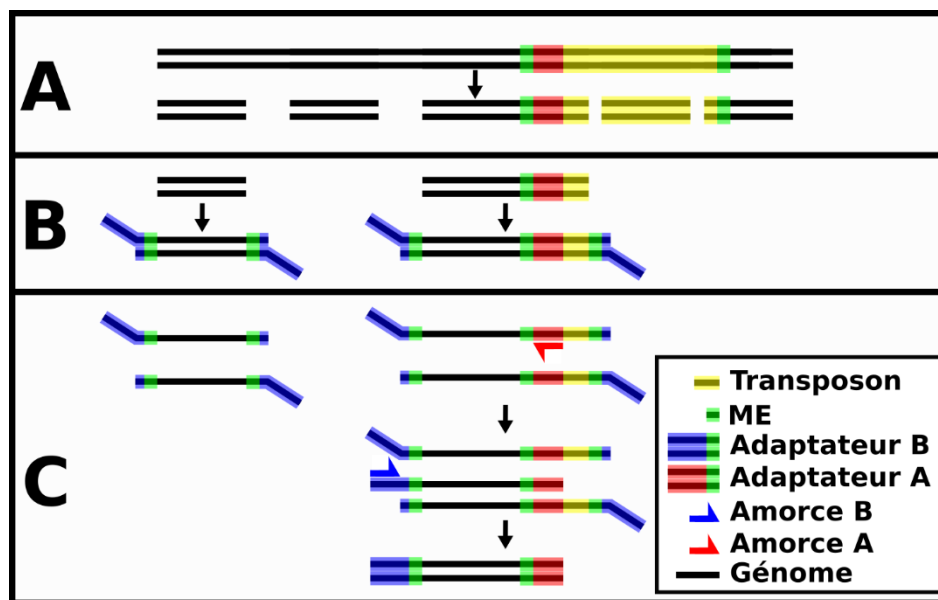


Figure 7. Préparation de la banque de séquençage des mutants d'insertion. (A) Extraction, fragmentation, réparation et adénylation de l'ADN génomique. (B) Ligation d'un adaptateur aux extrémités des fragments. (C) PCR pour obtenir les fragments prêts à être séquençés.

3.1.4 Analyses des mutants d'insertion

Seuls les échantillons pris aux jours 1, 2, 3 et 7 furent séquencés et seulement les trois derniers ont été pris en compte dans cette étude. Les cellules dans ces derniers échantillons, qui étaient suffisamment éloignés dans le temps pour ne plus contenir de mutants létaux ou de cellules donneuses, ont fait approximativement 24, 34 et 74 doublements, respectivement.

3.1.4.1 Saturation de la mutagenèse

La densité d'insertions des échantillons (entre $3,8 \times 10^1$ et $4,2 \times 10^1$ insertions/Kpb) fut considérablement moins élevée qu'attendu ($2,0 \times 10^3$ insertions/Kpb). Plusieurs étapes dans le processus menant aux résultats peuvent avoir causé cette baisse.

La première étant la mutagenèse générant les mutants. Bien que les tests d'UFC nous montraient que des millions de mutants avaient été générés, il serait possible que beaucoup de mutants fussent identiques, indiquant un biais d'insertion de la part de la transposase. Un motif d'insertion présentant une symétrie a en effet été observé avec nos données (figure 8) et correspond essentiellement à un motif observé dans une autre étude à partir de données d'insertion du transposon Tn5 *in vitro* (Kia *et al.*, 2017). Nos résultats supportent aussi la présence d'un certain biais vers les sites plus riches en GC mentionné par plusieurs (Green *et al.*, 2012 ; Kia *et al.*, 2017), mais celui-ci semble être mineur puisque la distribution du nombre de sites selon leur pourcentage GC suit globalement celle attendue pour le génome de DGF298 (figure 9A). Seuls les extrêmes, particulièrement les sites très pauvres en GC, étaient sous-représentés (figure 9B).

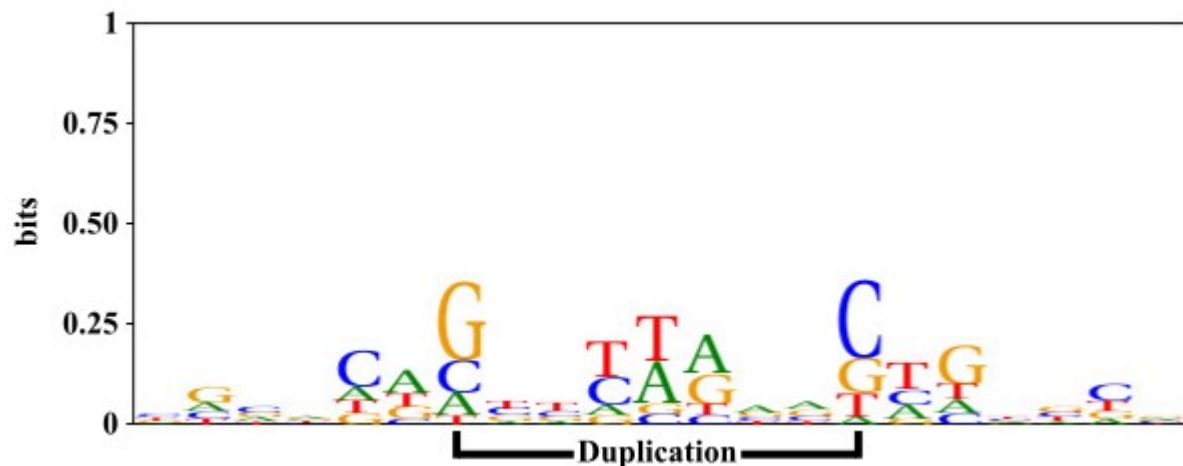


Figure 8. Motif d'insertion de la transposase Tn5. Motif d'insertion de 21 pb centré sur la séquence de 9 pb qui est dupliquée lors de l'insertion.

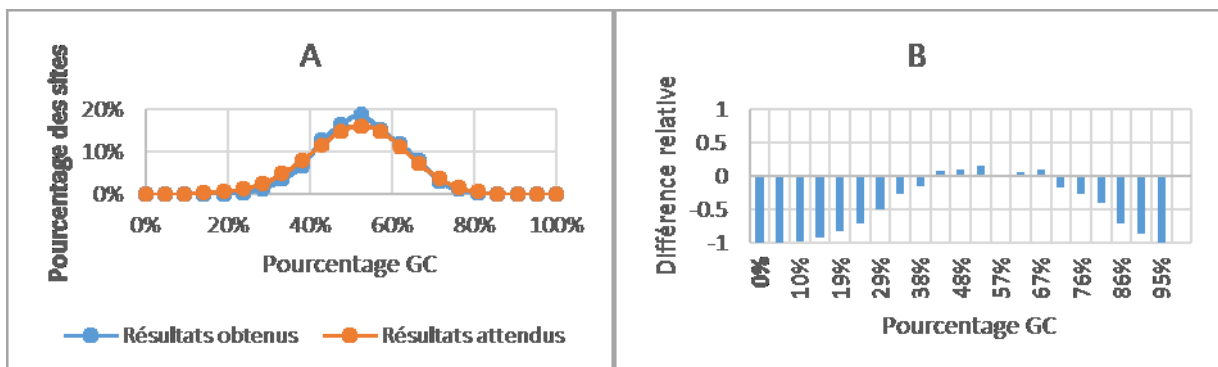


Figure 9. Taux d'insertion selon le contenu GC. (A) Distribution du pourcentage de sites selon leur contenu GC. La courbe bleue représente les sites d'insertion de 21 pb obtenus dans cette étude et la courbe orange, des séquences de 21 pb générées à partir du génome de DGF298. (B) Différence relative par rapport au nombre de sites attendus selon leur contenu GC.

Ces biais ont évidemment diminué la densité d'insertions, mais il est difficile de quantifier exactement l'impact qu'ils ont eu. Nous pourrions par contre nous servir de gènes comme *cmr*, rendant la cellule sensible à la spectinomycine, pour estimer le nombre de sites d'insertion uniques présents dans le mélange initial de mutants (Bohn and Boulloc, 1998). En effet, le nombre de cellules ayant ce gène interrompu a augmenté du jour 2 au jour 7, passant de 0,4 à

4,9 % de la population. Sa densité de transposons aussi a augmentée, étant à $2,9 \times 10^2$ insertions/Kpb au jour 2 et atteignant $4,7 \times 10^2$ insertions/Kpb au jour 7. Ce résultat indique que la densité d'insertions initiale du génome était plus élevée en termes de mutants différents, puisque de nouveaux sites sont découverts en analysant des échantillons plus tardifs. Il serait possible que les mutants correspondant à ces nouveaux sites fussent moins abondants initialement et qu'ils n'aient ainsi pas été séquencés. Leur abondance inférieure a pu être en partie causée par la différence entre la séquence de leur site d'insertion et celle du motif consensus.

Pour être sûr de ne pas attribuer trop hâtivement le manque de saturation au biais d'insertion, nous avons testé plusieurs autres facteurs. Un de ceux-ci était la profondeur de séquençage utilisée. Nous avons donc tracé les courbes du nombre de sites découverts en fonction du nombre de lectures séquencées. Ces courbes étaient logarithmiques et atteignaient leur plateau avec la quantité de séquences obtenues par échantillon, ce qui indique que le taux de séquençage était suffisant (figure 10).

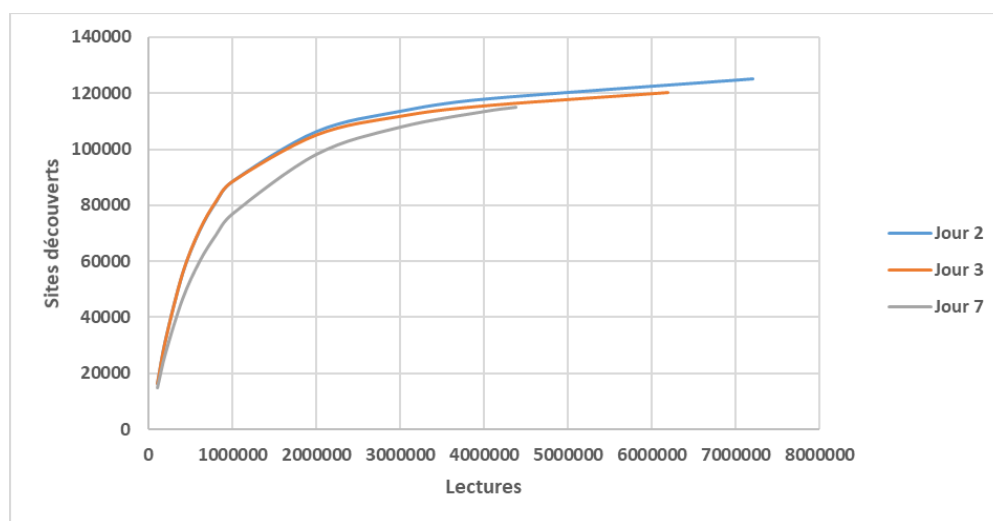


Figure 10. Saturation du séquençage. Nombre de sites d'insertion découverts en fonction du nombre de lectures pour les échantillons sélectionnés du suivi (le jour 2 en bleu, le jour 3 en orange et le jour 7 en gris).

Nous avons ensuite passé en revue la préparation de la banque de séquençage. Une réaction enzymatique peu efficace aurait en effet pu mener à des pertes aléatoires de mutants. Si c'était la cause de la baisse de la densité d'insertions, nous devrions pouvoir détecter de nouveaux mutants au cours du temps, puisque seulement une petite partie aléatoire des fragments atteindrait l'étape du séquençage. De plus, la couverture des nouvelles insertions détectées grâce aux temps plus tardifs devrait être similaire à celle des insertions détectées dans le premier temps, la sélection des fragments étant aléatoire. Bien que de nombreux sites nouveaux aient été découverts au cours du suivi (figure 11A), leur couverture moyenne était moins de la moitié de celle des sites récurrents (figure 11B). De plus, le motif d'insertion des nouveaux sites était plus dégénéré (figure 12). La préparation de la banque de séquençage ne semble donc pas être la cause principale du manque de saturation. Il est difficile de savoir exactement ce qui est arrivé, mais nous pourrions croire que certains mutants, initialement moins nombreux à cause du biais d'insertion, auraient, de manière aléatoire ou puisqu'ils sont devenus plus abondants, été inclus dans les échantillons plus tardifs du suivi.

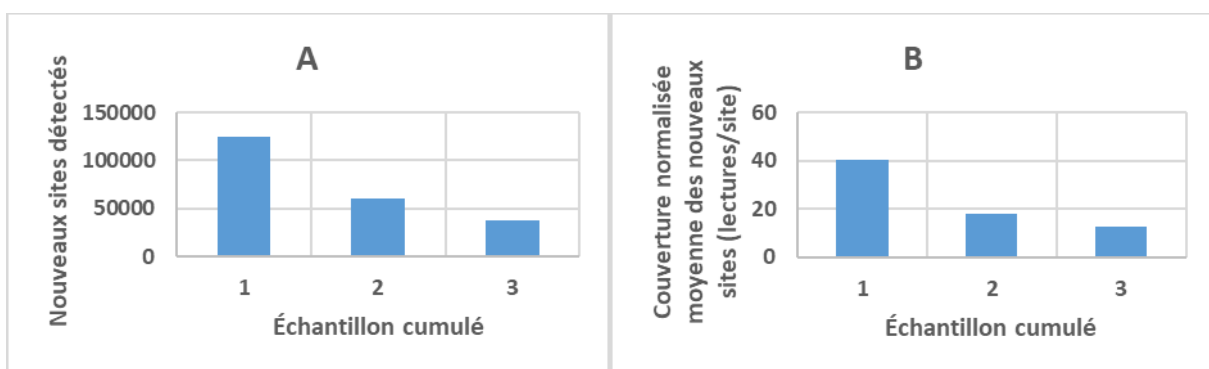


Figure 11. Nouveaux sites détectés au cours du suivi. (A) Le nombre de sites détectés à partir du premier échantillon sélectionné dans le suivi (jour 2) et le nombre nouveaux sites obtenus en cumulant séquentiellement les deux autres échantillons plus tardifs (jours 3 et 7). (B) Couverture normalisée (par le nombre de lectures alignées par échantillon) moyenne des sites détectés dans le premier échantillon sélectionné du suivi et celles des nouveaux sites obtenus en cumulant séquentiellement les deux autres échantillons plus tardifs.

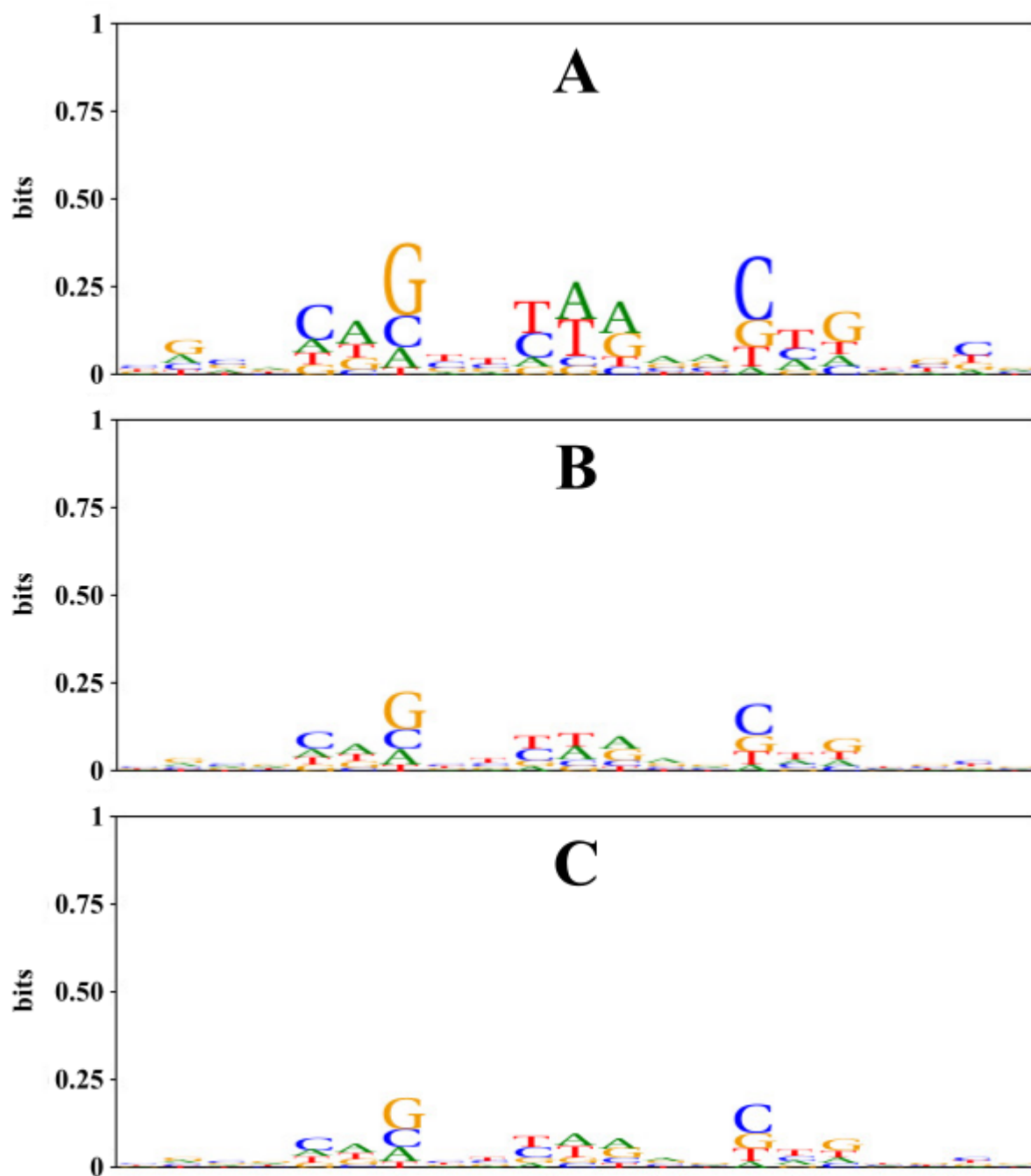


Figure 12. Motifs des sites détectés au cours du suivi. Motif des sites détectés dans le suivi avec le premier échantillon (A) et ceux générés avec les nouveaux sites découverts avec le deuxième (B) ou le troisième (C) échantillon.

En plus du biais d'insertion au niveau de la séquence, le niveau de couverture tout au long du génome semblait suivre un certain patron (figure 13). Il pourrait être le reflet de la distribution des gènes essentiels et/ou de l'accessibilité des sites au long du génome, dépendant de la séquence ou possiblement de la structure tridimensionnelle de celui-ci. Ce dernier effet pourrait donc s'ajouter aux différents biais diminuant la densité d'insertions globale du génome. Le niveau de couverture le plus élevé se situait systématiquement près de l'origine de réplication (~2,5 Mpb). La seule région montrant un grand écart entre les échantillons, au niveau du patron de la couverture, était celle contenant le gène *cmr* (~0,6 Mpb), puisque l'abondance des mutants de ce gène varie beaucoup au cours du suivi.

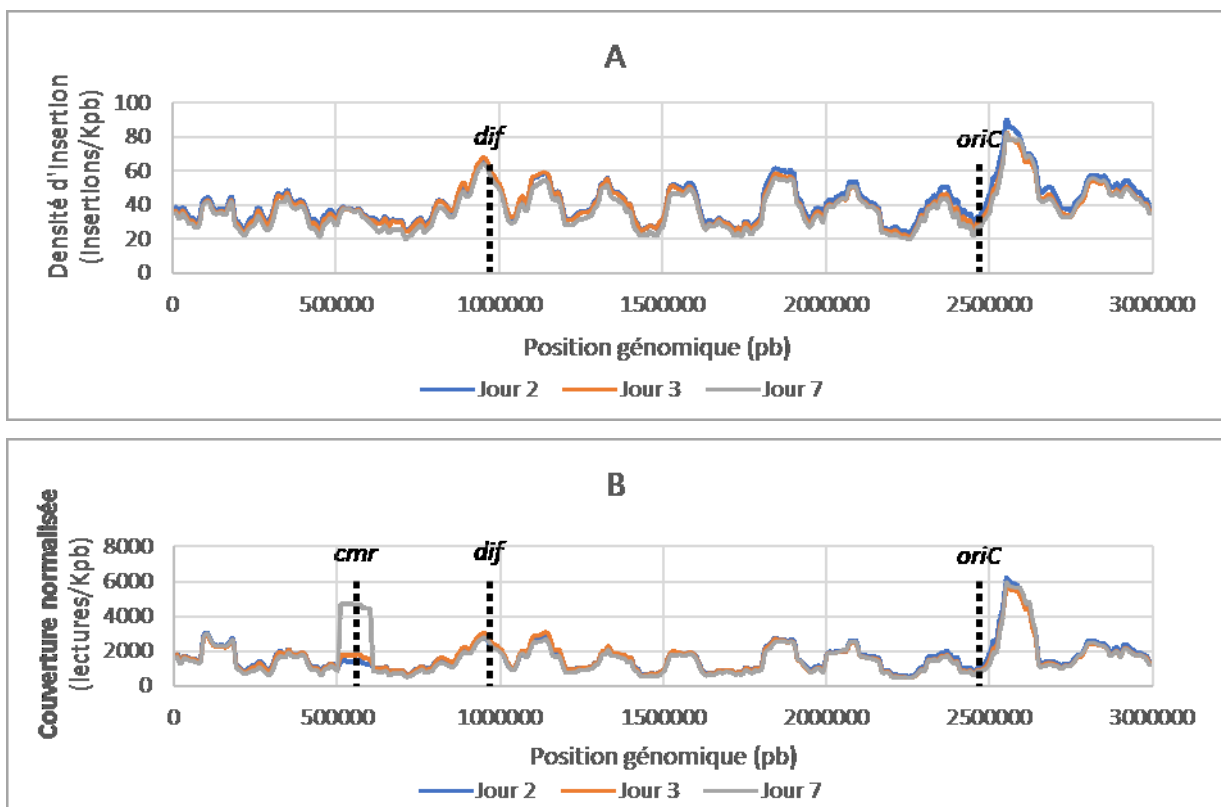


Figure 13. Localisation des sites d'insertion dans le génome de la souche DGF298. Densité d'insertions (A) et couverture normalisée (par le nombre de lectures alignées par échantillon) (B) moyennes du génome, calculées par régions de 100 Kpb avec chevauchement de 10 Kpb, pour les trois échantillons (le jour 2 en bleu, le jour 3 en orange et le jour 7 en gris).

3.1.4.2 Identification des gènes importants

Malgré la densité de transposons inférieure à nos attentes, les données ont permis d'identifier aisément la plupart des régions génomiques importantes (figure 14). Dans cet exemple, on peut bien voir que certains gènes avec des rôles importants comme *dnaA*, codant pour une protéine clé dans l'initiation de la réplication, contenaient peu d'insertions par rapport au reste du génome. Les sites d'insertions se trouvant dans de tels gènes correspondaient probablement à du bruit de séquençage qui sera traité dans la section suivante.



Figure 14. Exemple des résultats de la mutagenèse par insertions sur le UCSC genome browser. Les annotations bleues représentent les gènes ou les cicatrices laissées lors des délétions effectuées dans DGF298. Les petites lignes noires indiquent les sites d'insertion. Chaque rangée contenant ces lignes représente un des 3 échantillons sélectionnés, en ordre chronologique.

3.1.4.2.1 Bruits dans les données

Un contrôle sans transposon de la souche DGF298 a été ajouté à la banque de séquençage pour tenter de mesurer le nombre de lectures attribuées au mauvais échantillon. Nous avons ainsi pu estimer qu'il pouvait y avoir environ $2,3 \times 10^4$ lectures mal identifiées par échantillon, soit ~0,4 %. Les sites identifiés par celles-ci sont majoritairement (88,4 %) couverts d'une seule lecture, alors que la distribution des lectures chez les échantillons avec des transposons est plus étendue (tableau 4). Selon ces données, il serait donc bénéfique de retirer les sites couverts une seule

fois chez chacun des échantillons sélectionnés, puisque cela permettrait d'enlever 88,4 % du bruit pour une perte moyenne de seulement 10,6 % des sites totaux, en assumant que ce type de bruit se comporte de la même façon dans tous les échantillons.

Tableau 4. Bruits lors du séquençage

Couverture des sites (lectures/site)	Pourcentage des sites	
	Contrôle	Moyenne des 3 échantillons
1	88,4	10,6
2	6,9	4,8
3	1,7	3,8
4	0,8	3,4
5	0,6	3,1
≥6	1,8	74,3

Afin de mesurer l'impact d'un filtre sur l'identification des gènes essentiels, nous avons calculé la densité d'insertions moyenne, avec ou sans filtre, de 134 gènes classés comme essentiels dans plusieurs études distinctes par rapport à celle de tous les gènes (tableau 5; Baba *et al.*, 2006). En conservant seulement les sites avec plus d'une lecture, la densité de transposons des gènes essentiels diminue, de façon relative, davantage que la densité d'insertions globale des gènes, ce qui pourrait permettre de mieux distinguer les gènes importants des gènes facultatifs. Nous n'avons pas voulu augmenter le seuil minimal à plus d'une lecture par site pour éviter de trop diminuer la densité de transposons globale des gènes.

Tableau 5. Filtre de couverture minimale des sites d'insertion

Nombre de lectures minimal par site	Densité d'insertions des gènes (insertions/Kpb)	
	134 essentiels ¹	Tous
0	1,35	45,40
1	0,67	40,56
2	0,59	38,41
3	0,54	36,71

1. Ensemble de gènes tiré du troisième tableau supplémentaire de Baba *et al.* (2006) classés essentiels dans quatre études (Gerdes *et al.*, 2003 ; Kang *et al.*, 2004 ; Hashimoto *et al.*, 2005 ; Baba *et al.*, 2006).

3.1.4.3 Identification des gènes importants avec le *TraDIS toolkit*

Bien qu'il soit possible d'analyser les données de mutagenèse par transposons manuellement pour déduire le niveau d'essentialité d'un élément, plusieurs algorithmes ont été conçus pour automatiser le processus, ce qui permet, en plus du gain d'efficacité, d'appliquer strictement les mêmes paramètres de sélection à tous les gènes (Zomer *et al.*, 2012 ; Pritchard *et al.*, 2014 ; Solaimanpour *et al.*, 2015 ; Barquist *et al.*, 2016 ; Liu *et al.*, 2016 ; Yang *et al.*, 2017). Celui que nous avons partiellement utilisé, nommé le *Tradis toolkit*, est expliqué par Barquist *et al.* (2016). Nous avons privilégié leur approche, puisque, contrairement à d'autres, elle peut être utilisée pour une mutagenèse faite avec la transposase Tn5, elle ne requiert pas de connaissances a priori sur l'essentialité des gènes et elle ne prend pas en compte la couverture des différents sites qui, comme vu dans les sections précédentes, est biaisée. Brièvement, il a d'abord fallu calculer un indice d'insertion pour chacun des gènes en normalisant le nombre de sites d'insertion dans le gène par sa longueur. Ensuite, nous nous sommes servis du script R *tradis_essentiality* pour premièrement créer une distribution groupant les gènes selon leur indice d'insertion. Celle-ci

possédait deux modes, un correspondant aux gènes essentiels, avec des indices d'insertion faibles, et un aux gènes non essentiels, avec des indices d'insertion plus élevés. L'algorithme a ensuite adapté une distribution gamma à chacun des modes et a calculé une probabilité d'appartenance à chacune des distributions. Les gènes qui étaient situés entre les deux modes et qui ne permettaient pas d'être significativement classés furent identifiés comme étant ambigus.

Parmi les 2678 séquences codantes prises en compte dans l'analyse, plus de 75 % (2017) ont été classées essentielles dans aucun des échantillons (tableau supplémentaire 1). Le tableau 6 montre le nombre de gènes en fonction du nombre d'échantillons supportant leur essentialité. Selon ces résultats, il pourrait être possible de retirer, en cumulant la longueur des séquences codantes facultatives, plus de 1,9 Mpb du génome.

Tableau 6. Résumé de l'analyse TraDIS

Nombre d'échantillons appuyant l'essentialité	Nombre de gènes
0	2017
1	107
2	86
3	468

Bien que l'interruption individuelle de ces 2017 gènes ne semble pas affecter assez la croissance pour qu'ils soient classés essentiels après 73 doublements, il serait possible que certains d'entre eux aient tout de même eu un impact sur le taux de croissance. En comparant la couverture des gènes entre les premiers et les derniers temps, nous avons pu observer que l'interruption de 46 gènes, parmi ces 2017, pouvait réduire le taux de croissance relatif des cellules à moins de 95 % (figure 15 ; voir section 2.9.1 pour la formule utilisée ; van Opijnen *et al.*, 2009). Ainsi,

seulement 707 séquences codantes seraient importantes pour avoir une bonne croissance dans les conditions testées.

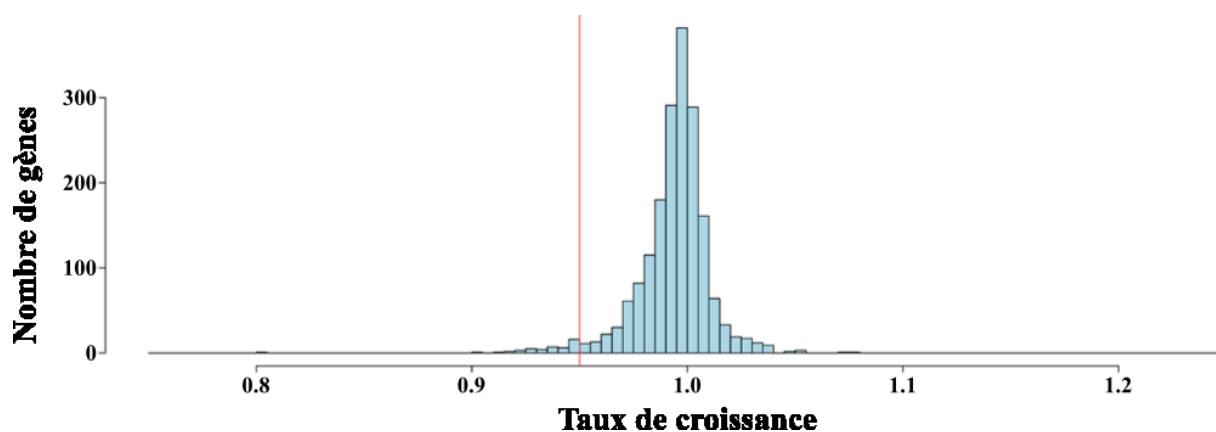


Figure 15. Taux de croissance relatif des mutants de gènes classés facultatifs selon TraDIS. Distribution des gènes selon le taux de croissance de leurs mutants. Les 46 gènes dont leurs mutants avaient un taux de croissance inférieur à 95 % (indiqué par la ligne rouge), ont été ajoutés à la liste de gènes importants identifiés par le *TraDIS toolkit* (Barquist et al., 2016).

Nous nous sommes ensuite intéressés aux fonctions des 1971 gènes ne semblant pas être importants pour la croissance. Plus de 80% des gènes liés aux fonctions telles que la sécrétion et la motilité cellulaire, le transport et le métabolisme des carbohydrates, la biosynthèse, le transport et le catabolisme de métabolites secondaires et aux fonctions inconnues ou seulement prédites sont facultatifs (figure 16). En regardant plus en détails les fonctions de ces séquences codantes facultatives, nous pouvons voir que beaucoup d'entre elles sont liées à la survie en conditions spécifiques, telles qu'en absence d'oxygène (anaérobie), avec une source de carbone autre que le glucose, en milieu minimal, avec une carence en certains éléments, en présence d'antibiotiques ou de métaux toxiques, en milieu acide, à température élevée ou encore sous stress osmotique (tableau supplémentaire 1).

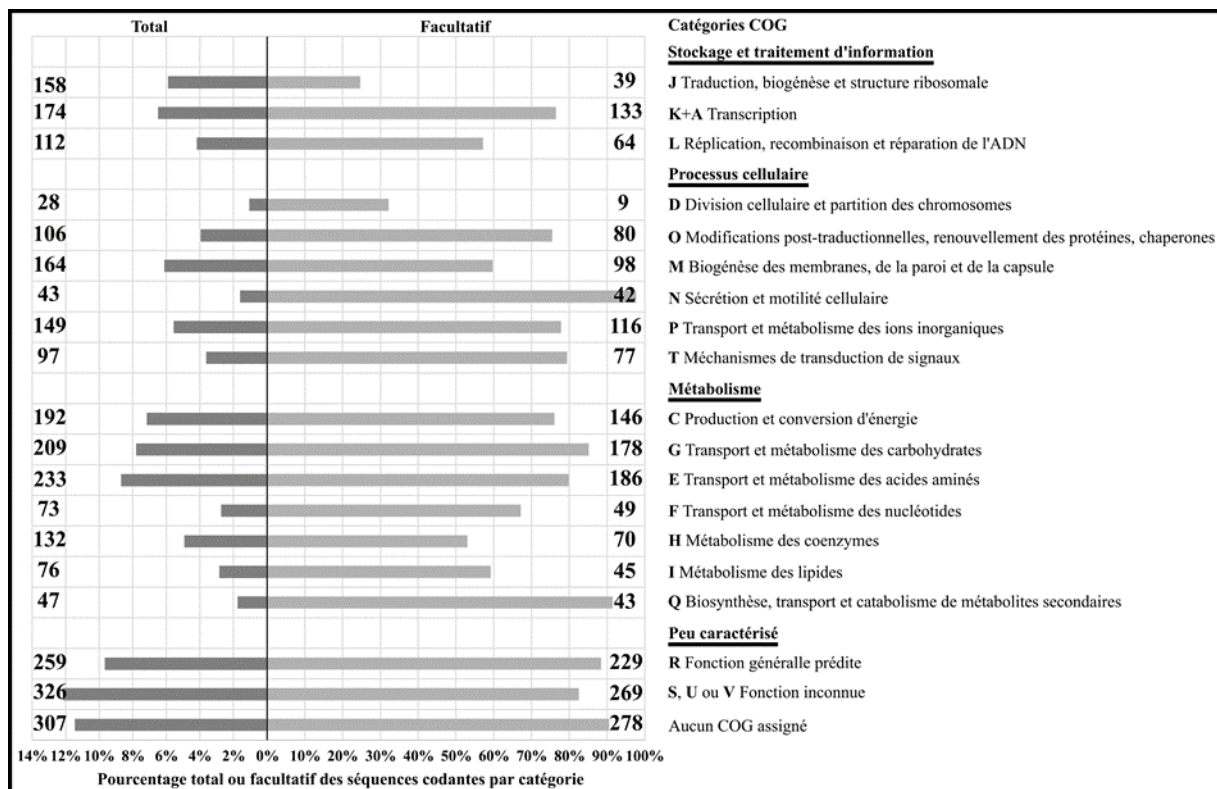


Figure 16. Fonctions des séquences codantes facultatives à DGF298. Pourcentage des séquences codantes totales appartenant à une certaine catégorie et pourcentage des séquences codantes étant facultatives dans ces catégories.

3.2 Mutagénèse par délétions

3.2.1 Génération des mutants de délétion

Bien que les données de mutagénèse par insertions soient très informatives sur l'essentialité des gènes, seul le retrait de ceux-ci peut réellement confirmer que le génome de DGF298 peut être réduit davantage. Pour cette raison, nous avons aussi fait une mutagénèse par délétions avec cette souche. Tel qu'illustré en figure 17, nous avons mis au point une méthode, inspirée de Goryshin *et al.* (2003), permettant de générer des mutants de délétion aléatoire. De façon très similaire à la méthode que nous avons utilisée pour effectuer les mutagénèses par insertion, un transposon a été inséré dans le génome de la cellule (Fig. 17A). Par contre, cette méthode utilise le transposon du plasmide pFG058 qui est plus complexe que celui du plasmide pFG051. En effet, ce transposon contenait un gène codant pour une transposase différente de la première, de type mariner (Robertson and Lampe, 1995). L'expression de celle-ci était inhibée en présence de fer grâce au promoteur P_{fhuA} réprimé par la protéine Fur, endogène à *E. coli* (Guan *et al.*, 2013). Après avoir sélectionné les mutants d'insertions, nous avons donc ajouté un chélateur pour réduire la concentration cellulaire en fer et ainsi induire la production de la transposase. Celle-ci a ensuite reconnu deux sites spécifiques situés dans le transposon et a alors formé un transposome (Fig. 17B). Dû à l'orientation de ces sites, le génome entier était alors utilisé comme transposon. Le seul endroit où ce dernier pouvait être inséré était dans le génome bactérien, ce qui pouvait, entre autres, causer la fission du chromosome (Fig. 17C-D). Une des deux structures obtenues n'avait alors pas d'origine de réplication et était donc perdue au fil des divisions, si la délétion n'était pas létale. Après l'induction, le plasmide pFG057 codant pour l'endonucléase Cas9 et pour un ARNg ciblant une région du transposon fut introduit dans la cellule par conjugaison, ce qui devait causer la mort de celles n'ayant pas effectué de délétion (Fig. 17E ; Mali *et al.*, 2013). Un suivi par passages de 7 jours dans le milieu riche et défini a ensuite été fait avec les mutants de délétion.

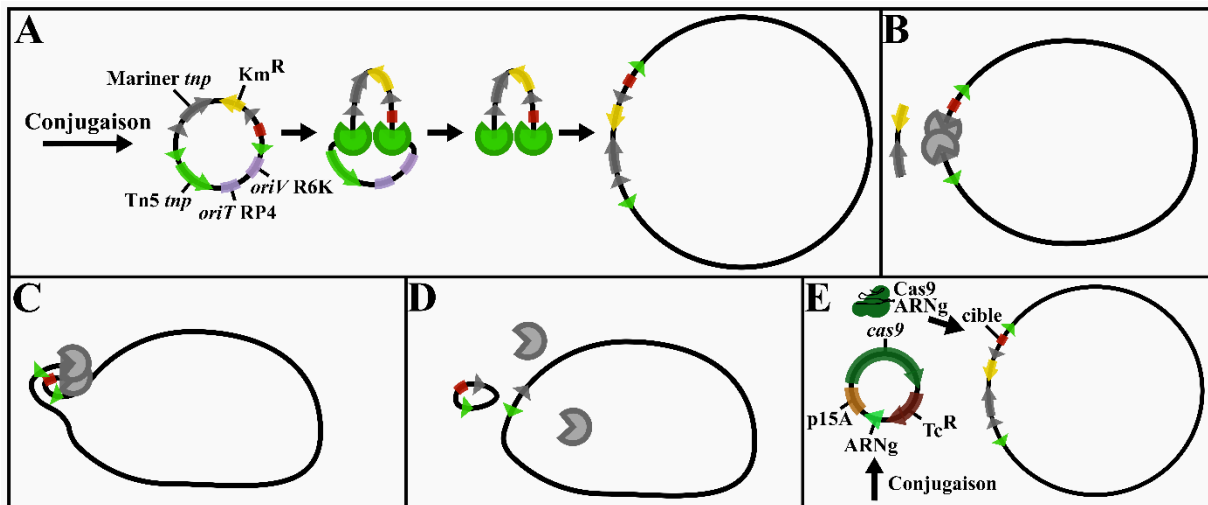


Figure 17. Mutagenèse par délétions aléatoires. (A) insertion d'un transposon spécialisé dans le génome. (B) Formation du transposome suite à la production de la seconde transposase. (C-D) Transposition et fission du chromosome. (E) Contre-sélection par conjugaison d'un plasmide codant pour le complexe Cas9-ARNg qui cible une partie du transposon absente chez les cellules ayant fait la délétion voulue.

3.2.2 Confirmation des mutants de délétion

Plusieurs confirmations ont été effectuées avec les mutants de délétion. D'abord, des tests d'UFC ont été faits pour valider leur sélection lors du suivi (figure 18). Les combinaisons d'antibiotiques choisies ont permis de suivre l'évolution de quatre groupes de cellules, soit celles ayant fait ou non la seconde transposition et ayant reçu ou non le plasmide pFG057. Il était attendu que les cellules non induites portant encore le transposon entier soient résistantes à la kanamycine, et, une fois l'induction faite, qu'elles perdent cette résistance et acquièrent celle à la tétracycline grâce au plasmide de contre-sélection pFG057. Puisque le suivi était fait en milieu contenant de la tétracycline, seules les cellules ayant pu maintenir ce plasmide, qui avaient donc une délétion, devaient croître. Nos données indiquent que c'était majoritairement le cas, mais qu'il y avait une portion non négligeable, mais qui déclinait au cours du temps, de cellules qui portaient encore le transposon.

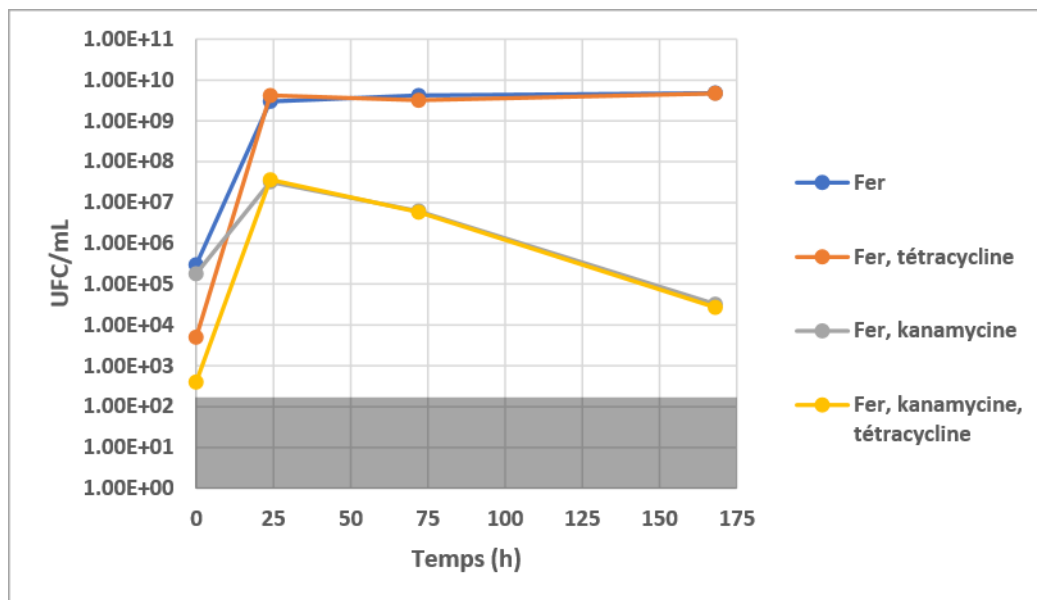


Figure 18. Confirmation de la sélection des mutants par compte d'UFC. Quantité des différentes cellules de la souche DGF298 au cours du suivi, soit toutes les cellules viables (bleu), celles avec pFG057 (tétracycline ; orange), celles contenant le transposon complet (kanamycine ; gris) et celles contenant le transposon complet et pFG057 (kanamycine et tétracycline ; jaune). La zone grise représente la limite de détection de la méthode de compte d'UFC.

Des tests additionnels par PCR ont été faits pour confirmer la nature des cellules sélectionnées en présence de tétracycline, avec ou sans kanamycine. 10 colonies provenant d'un échantillon pris au jour 7 du suivi ont été testées, 5 isolées d'une gélose qui contenait de la tétracycline et 5 autres d'une gélose avec de la tétracycline et de la kanamycine. Toutes celles testées qui ont crû en présence de tétracycline seulement semblaient être des mutants de délétions, puisque la partie droite du transposon ne pouvait être amplifiée (figure 19A). Pour ce qui des colonies qui pouvaient croître aussi avec de la kanamycine, elles semblaient contenir le transposon complet puisque les régions gauche et droite ont pu être amplifiées. De plus, un autre test PCR a permis de confirmer que la résistance à la tétracycline de ces dernières cellules était bien due au plasmide pFG057, ce qui indique que leur système de contre-sélection était défaillant (figure 19B).

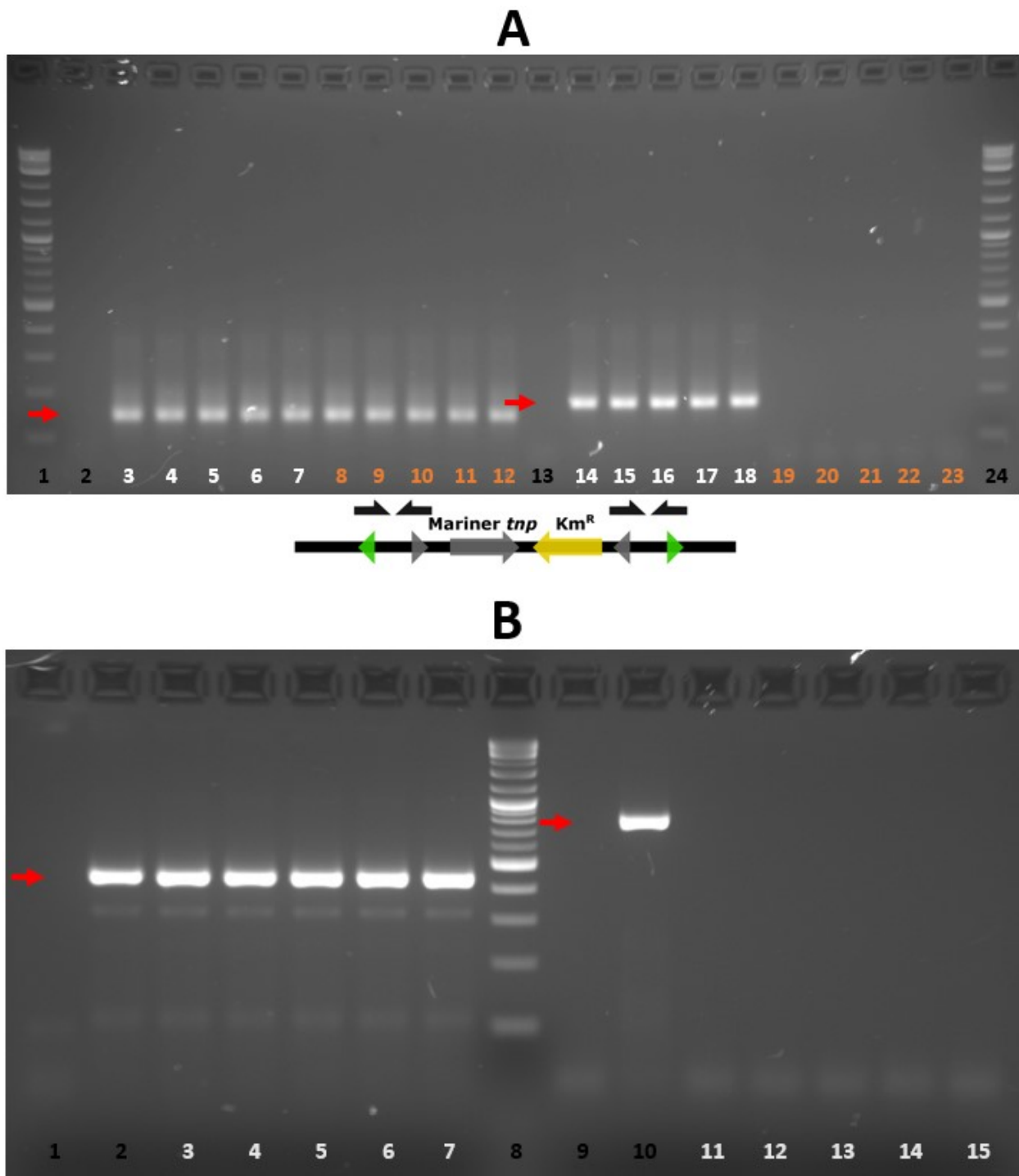


Figure 19. Confirmation des mutants de délétion par PCR. (A) Gel électrophorèse avec les produits d'amplification des régions gauche (150 pb) et droite (177 pb) du transposon. Les chiffres blancs désignent les 5 colonies qui provenaient d'un échantillon pris au jour 7 et qui ont crues sur une gélose avec tétracycline et kanamycine. Les chiffres oranges désignent 5 colonies

qui provenaient d'un échantillon pris au jour 7 et qui ont cru sur une gélose avec tétracycline seulement. (B) Gel d'électrophorèse avec les produits d'amplification d'une région présente dans le plasmide pFG057 (à gauche de l'échelle moléculaire ; 440 pb) ou dans le plasmide pFG036 (à droite de l'échelle moléculaire ; ~750 pb). Les chiffres blancs désignent les 5 colonies qui provenaient d'un échantillon pris au jour 7 et qui ont crû sur une gélose avec tétracycline et kanamycine. Les plasmides purifiés ont servi de matrice pour les contrôles positifs (puits 2 et 10). Un contrôle négatif contenant de l'ADN génomique de la souche DGF298 a été inclus pour chacun des produits (puits 2 et 13 en A et puits 1 et 9 en B). Les produits attendus sont indiqués par les flèches rouges.

Pour cette mutagenèse, il était difficile d'estimer le nombre de mutants différents générés puisque le moment exact de leur création n'était pas connu, l'induction s'étant échelonnée sur 8 h à 30°C. Nous pourrions cependant situer le nombre de mutants différents maximal avec les tests d'UFC et obtenir une borne inférieure en ajustant ce nombre avec le nombre de doublements approximatif effectués pendant l'induction. Ainsi, entre $1,1 \times 10^5$ et $1,6 \times 10^6$ mutants de délétions auraient été générés, en assumant que ceux-ci aient été créés au même moment, en quantité égale et qu'ils aient la même vitesse de croissance.

3.2.3 Séquençage des mutants de délétion

Du séquençage haut débit multiplexé en paire a été utilisé pour identifier la position du premier et du second évènement de transposition chez les mutants présents dans les échantillons pris au cours du suivi. Pour ce faire, nous avons mis à profit la portion du transposon, portant l'adaptateur Nextera A, qui reste dans le génome suite à la délétion (figure 20). Lors de la préparation de la banque de séquençage, l'ADN a d'abord été fragmenté et réparé, puis une ligation intramoléculaire fut effectuée (figure 20A-B). Cette étape permettait d'obtenir des molécules circulaires permettant, à l'aide d'une amplification PCR, d'obtenir un produit linéaire portant à chacune de ses extrémités les régions qui se trouvaient respectivement en amont et en aval de la cicatrice produite lors de l'évènement de délétion. Les deux réactions PCR ont donc été faites, ce qui a permis de générer des fragments portant les deux adaptateurs Nextera, A et

B (figure 20C). Chaque paire de lectures produite, suite au séquençage Illumina en paire, permet ainsi d'identifier les deux zones bordant la délétion.

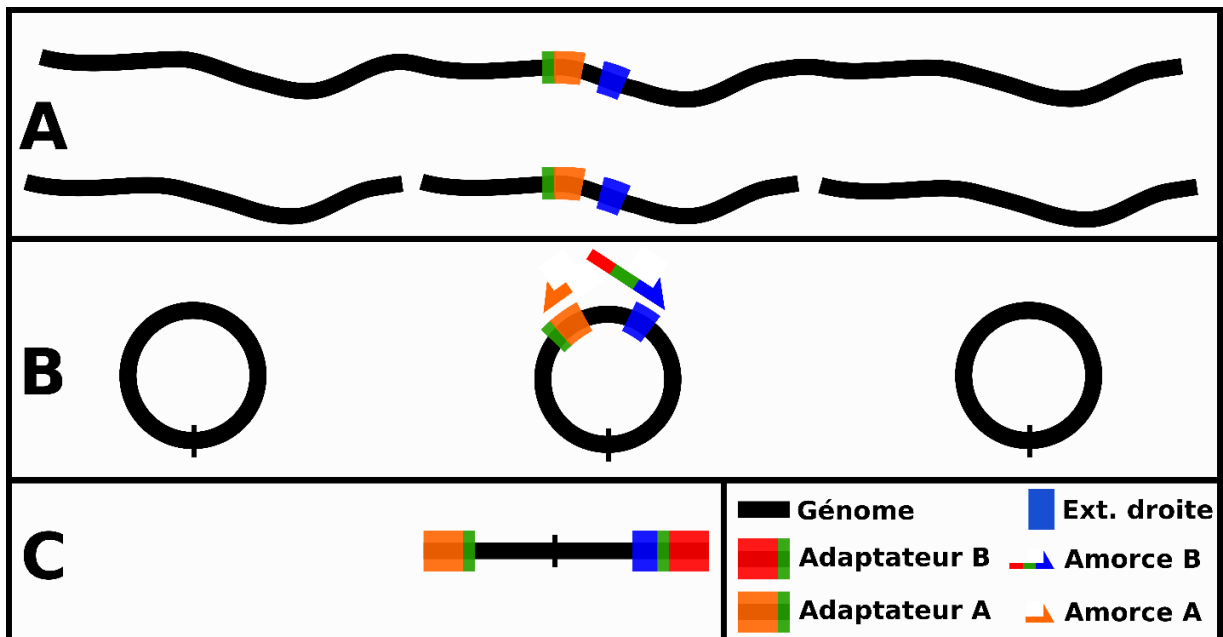


Figure 20. Préparation de la banque de séquençage des mutants de délétion. (A) Extraction, fragmentation et réparation de l'ADN génomique. (B) Ligation intramoléculaire et orientation des amorces PCR pour obtenir le produit à séquençer. (C) Structure du produit d'amplification prêt pour le séquençage.

3.2.4 Analyses des mutants de délétion

Seuls les échantillons du jour 2 au jour 7 ont été pris en compte dans la suite des analyses pour éviter les délétions létales pouvant être présentes dans les premiers temps.

3.2.4.1 Saturation de la mutagenèse

Le nombre de mutants de délétion différents détectés dans les échantillons des jours 2 à 7 ($8,8 \times 10^2$) fut moins élevé qu'attendu ($\geq 1,1 \times 10^5$). Alors que le biais d'insertion peut avoir été la principale cause du manque de saturation pour la mutagenèse par insertions, il est peu probable que ce soit encore le cas pour cette mutagenèse. Ceci puisque la transposase mariner requérait seulement un site TA pour l'insertion et que le motif de la transposase Tn5 était suffisamment dégénéré pour générer plus de 100000 mutants distincts selon nos résultats précédents.

Comme pour la mutagenèse par insertions, un fort biais de sélection pour les mutants d'un certain gène pourrait être utilisé pour estimer le nombre de délétions uniques initiales. En effet, les cellules qui portaient une délétion du gène *ompR*, lié à la diffusion à travers la membrane externe, semblaient avoir un avantage de croissance puisque le nombre de délétions uniques l'affectant a augmenté au cours du temps, atteignant jusqu'à 40 délétions dans un même échantillon et s'accaparant 99,8% des lectures correspondant à des délétions (figure 21). Ces 40 délétions couvraient ensemble l'entièreté d'une région de 13,7 Kpb. En extrapolant sur le génome, nous pourrions donc estimer qu'il y avait plus de 8000 délétions différentes parmi les mutants initiaux. C'est toujours moins que notre estimation initiale, mais c'est environ 10 fois plus que le nombre de délétions détectées pour les jours 2 à 7. Nous avons donc continué à chercher la raison pour laquelle nous n'avons pas détecté toutes ces délétions.

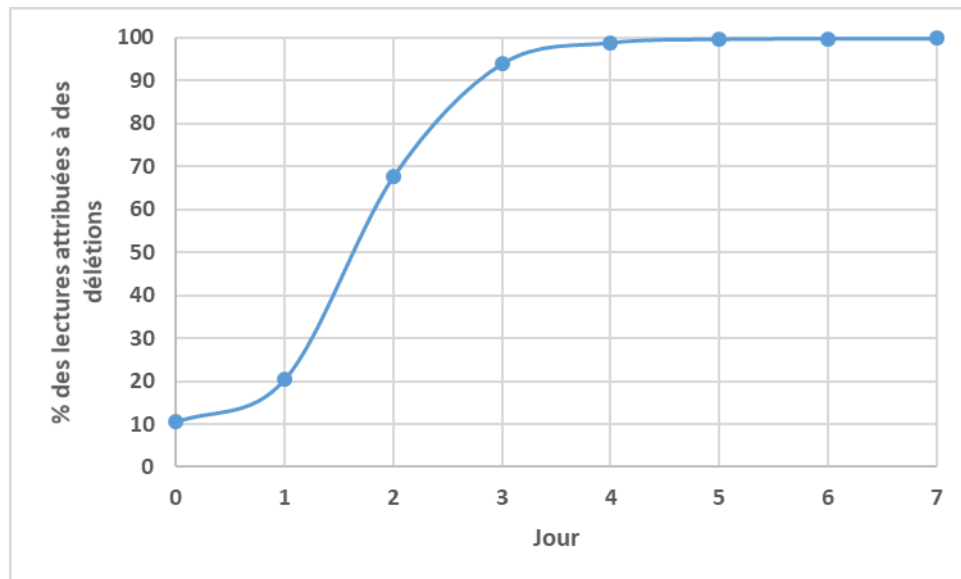


Figure 21. Abondance relative des délétions affectant le locus *ompR*. Pourcentage des lectures attribuées à des délétions dans une région de 13,7 Kpb autour du gène *ompR*.

Le problème ne semblait pas être au niveau de la profondeur de séquençage, puisque les courbes de saturation atteignaient rapidement leur plateau (figure 22). On peut également voir dans cette dernière figure que le nombre de délétions uniques diminue au cours du temps, ce qui coïncide avec la montée du nombre de mutants d'*ompR* (figure 21).

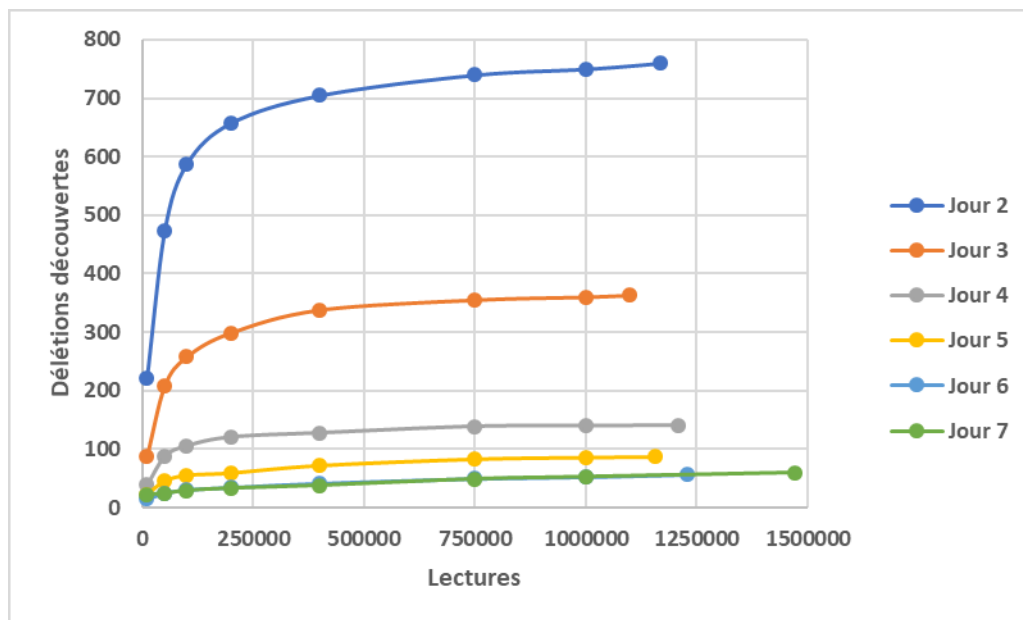


Figure 22. Saturation du séquençage de la mutagenèse par délétions. Nombre de délétions découvertes en fonction du nombre de paires de lectures séquencées.

Nous avons ensuite investigué la préparation de la banque de séquençage. La méthode utilisée, impliquant une étape de ligation intramoléculaire, a pu contribuer à la diminution du nombre de mutants détectés. Effectivement, en moyenne, pour les échantillons des jours 2 à 7, 16 % des lectures obtenues correspondaient à des produits de ligation ayant au moins une des deux régions génomiques bordant la cicatrice de taille inférieure à 30 pb. Ces lectures ont été interprétées comme de mauvais alignements ou comme une délétion de taille similaire à celle du génome, dans le cas où l'une des régions est absente (figure 23).

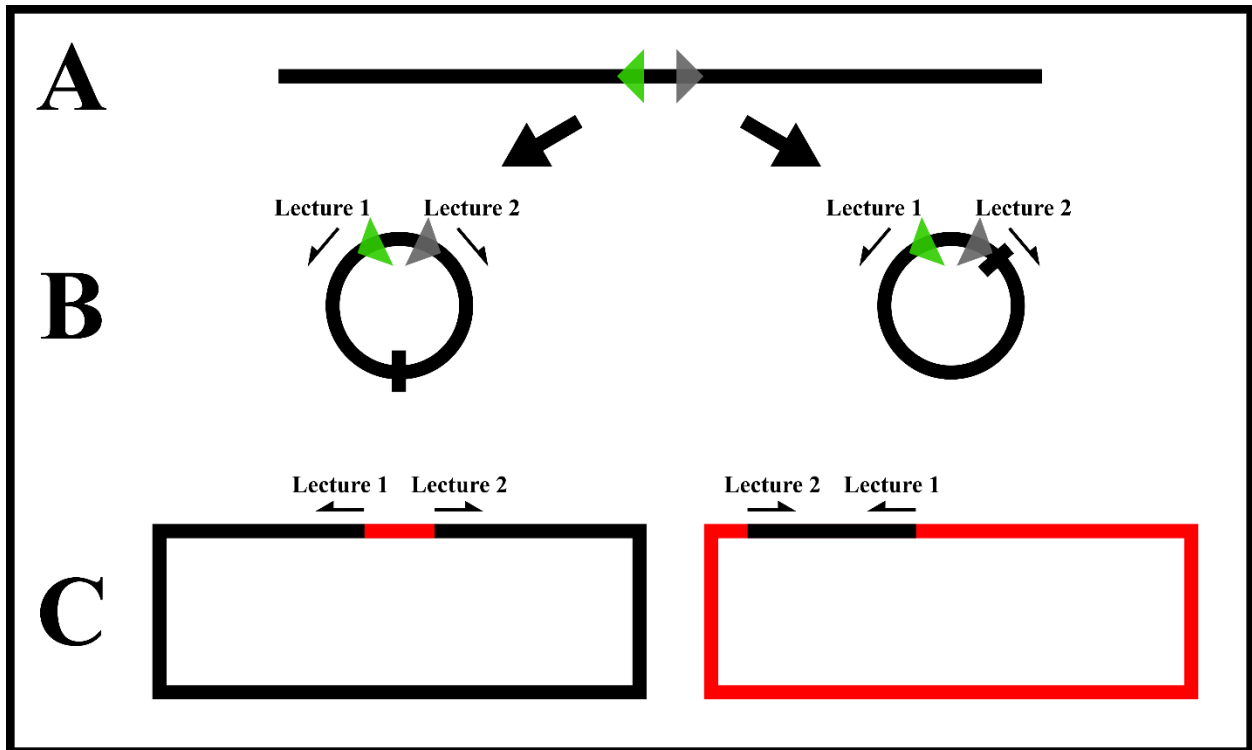


Figure 23. Artéfacts générés durant la préparation de la librairie de séquençage. (A) ADN génomique contenant une cicatrice de délétion. (B) 2 produits distincts obtenus suite à la fragmentation et à la ligation de l'ADN du même mutant initial. (C) Interprétation des résultats de séquençage, où les délétions sont indiquées en rouge.

Outre ces artéfacts, une faible efficacité au niveau de la ligation, ou autre réaction enzymatique, aurait pu mener à des pertes aléatoires de mutants. Si c'était la cause du manque de saturation, nous devrions pouvoir détecter de nouveaux mutants au cours du temps puisque seulement une petite partie aléatoire des fragments atteindrait l'étape du séquençage. De plus, la couverture des nouvelles délétions détectées grâce aux temps plus tardifs devrait être similaire à celle des délétions détectées dans le premier temps, la sélection des fragments étant aléatoire. Bien que de nouvelles délétions aient été découvertes au cours du suivi (figure 24A), leur couverture était inférieure (figure 24B). La préparation de la banque de séquençage ne semble donc pas être la cause du manque de saturation.

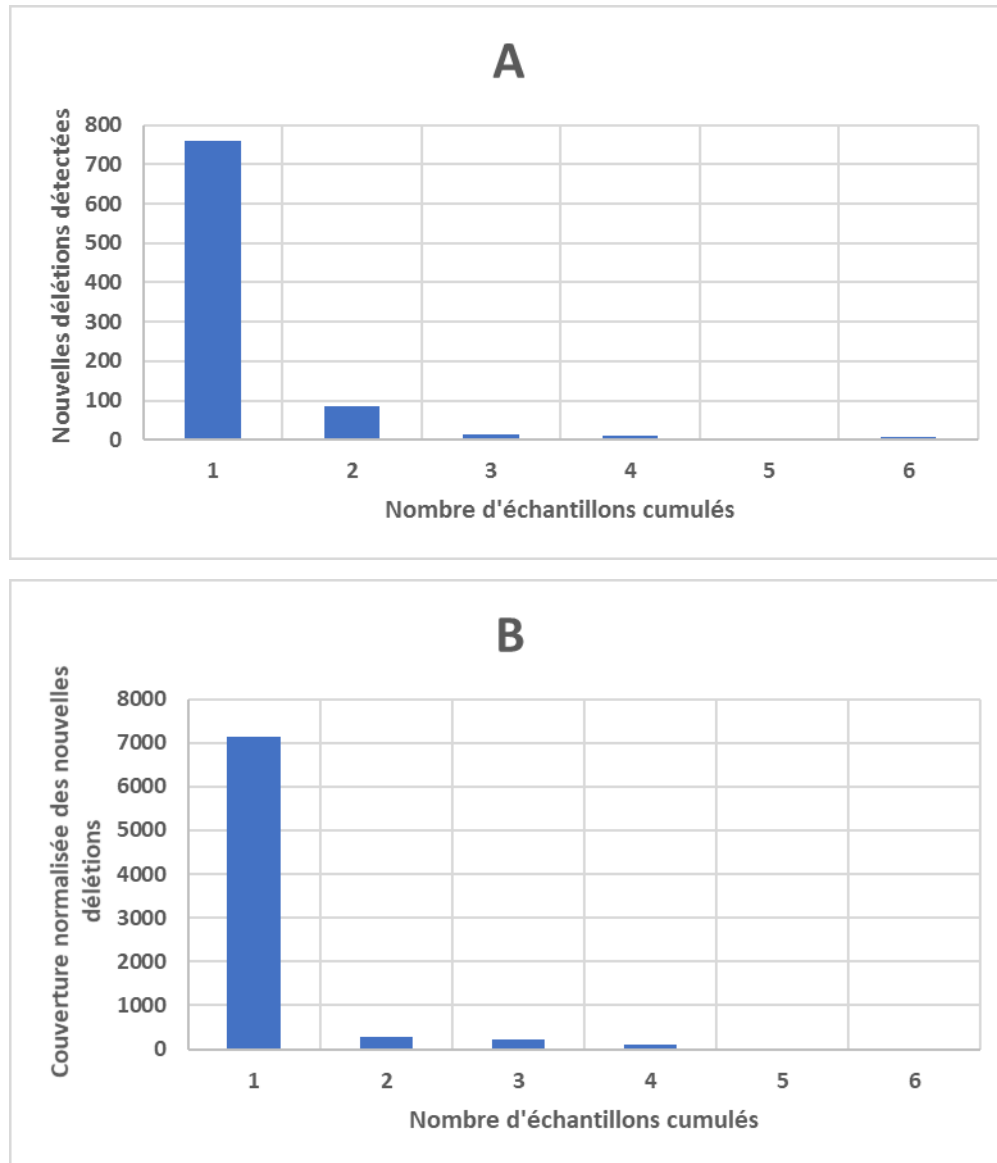


Figure 24. Nouvelles délétions détectées au cours du suivi. (A) Le nombre de délétions détectées avec le premier échantillon sélectionné du suivi (jour 2) et le nombre nouvelles délétions obtenues en cumulant séquentiellement cinq autres échantillons plus tardifs (jour 3 à 7). (B) Couverture normalisée des délétions détectées dans le premier échantillon sélectionné du suivi et celles des nouvelles délétions obtenues en cumulant séquentiellement cinq autres échantillons plus tardifs.

Nous avons aussi détecté une forte présence de cellules portant une fusion de pFG058 et pFG036. En effet, cette fusion a été créée dans la cellule donneuse portant les deux plasmides

et fut causée par la fuite d'expression du gène de la transposase mariner, qui a reconnu ses sites dans le transposon de pFG058, a formé un transpososome avec l'intégralité du plasmide, à l'exception des gènes codant pour la transposase et la résistance à la kanamycine, et l'a inséré dans pFG036. Il s'agissait alors d'un plasmide qui possédait une origine de transfert, deux origines de répllication dont une permettant au plasmide de se répliquer aussi dans la cellule receveuse, un gène de résistance à la tétracycline, un gène codant pour la transposase Tn5 et le gène du répresseur cI inhibant l'expression cette dernière. En moyenne, 33,5 % des lectures obtenues pour les échantillons des jours 2 à 7 correspondaient à ce produit (figure 25 ; lecture 1 s'alignant sur pFG058 et lecture 2 sur pFG036).

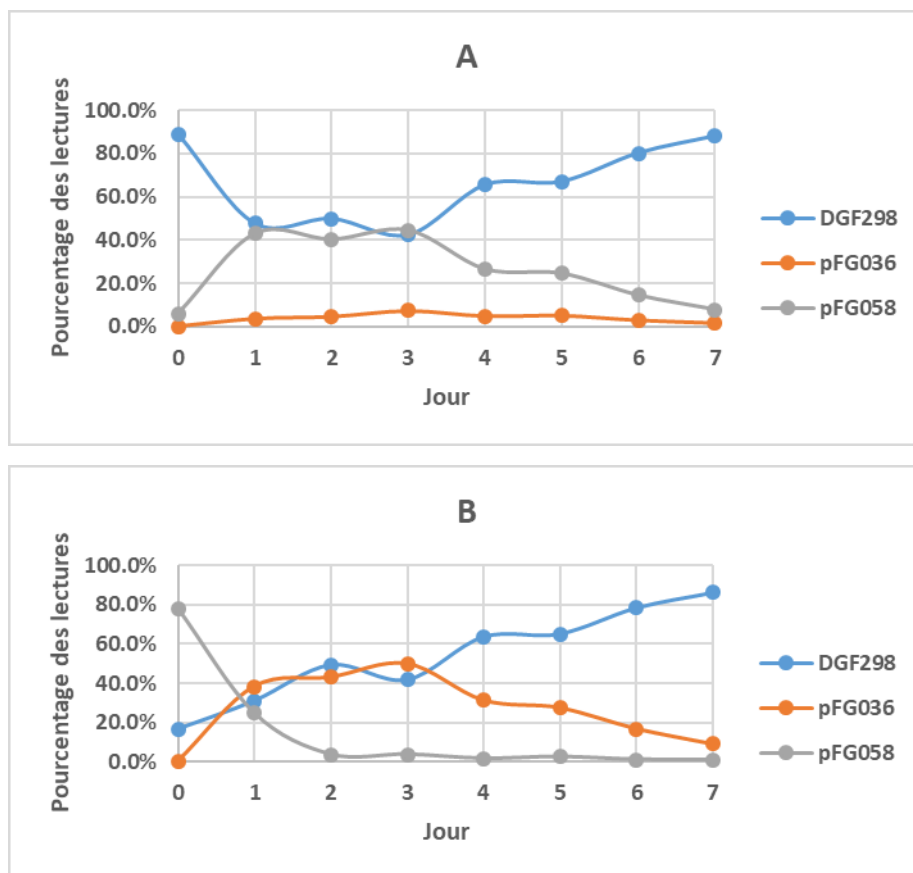


Figure 25. Alignement des lectures. Pourcentage des lectures 1 (A) et 2 (B) s'alignant sur le génome de DGF298 (bleu), le plasmide pFG036 (orange) ou le plasmide pFG058 (gris) pour les 8 échantillons séquencés.

L'abondance relativement faible de ce produit au début du suivi suggère que c'est seulement par la suite, avec la sélection à la tétracycline, que les cellules le possédant sont devenues abondantes. Leur quantité a tout de même diminuée vers la fin du suivi face aux mutants de délétion du locus *ompR*. Même si la présence de ce produit n'était pas souhaitable, elle ne semble pas avoir été la cause du manque de saturation des événements dans le génome, puisque leur abondance initiale était faible et que le séquençage était malgré tout à saturation.

Finalement, nous avons analysé l'abondance relative des délétions détectées tout au long du génome (figure 26). Nous pouvons observer qu'il y avait un biais au niveau de l'origine de réplication (~2,5 Mpb) et à deux autres loci. Les gènes les plus affectés dans le premier locus (~0,3 Mpb) étaient ceux de l'opéron *cyoABCDE* et *ampG*. Le gène le plus touché dans le second locus (~2,7 Mpb) était *ompR*, qui, comme vu précédemment, était de plus en plus muté au cours du temps. Le biais pour les deux premières régions était présent même dans l'échantillon pris au tout début du suivi, ce qui suggère que les insertions étaient effectuées préférentiellement dans ces régions, ou encore que ces mutants d'insertion (avant l'induction de la deuxième transposition) ou de délétion avaient un avantage de croissance qui les a rendus prédominants dans la population avant le début du suivi, ce qui implique que l'estimation du nombre de délétions initiale basée sur *ompR* est trop élevée.

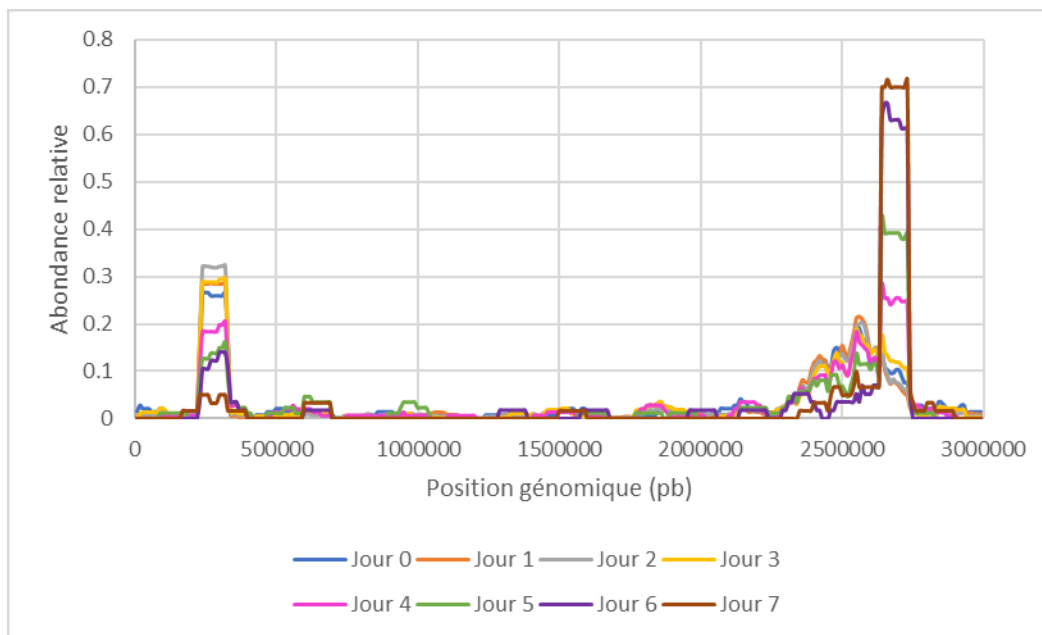


Figure 26. Localisation des délétions dans le génome de la souche DGF298. Nombre de délétions se trouvant dans des régions de 100 Kpb avec chevauchement de 10 Kpb relatif au nombre total de délétions détectées dans l'échantillon.

Nous pourrions conclure que le nombre de mutants initiaux différents devait être plus élevé, mais que certains étaient beaucoup plus nombreux, possiblement à cause d'un biais au niveau de la position dans le génome, à l'avantage de croissance de certains mutants et à un faible taux d'induction.

3.2.4.2 Sommaire de la mutagenèse

Parmi 2431 délétions identifiées initialement, 2110 se retrouvaient dans les jours 2 à 7. 118 de ces délétions qui devraient être viables affectaient au moins un des 134 gènes essentiels mentionnés précédemment (voir section 3.1.4.2.1) et 144 affectaient au moins un des 468 gènes classés importants dans les trois temps pris en compte dans la mutagenèse par transposons. Toutes ces délétions fautives avaient une taille supérieure à 50 Kpb et pourraient être d'autres

artéfacts de séquençage, en plus de ceux mentionnés précédemment, pouvant être causés par des ligations intermoléculaires lors de la préparation de la banque de séquençage. De plus toutes les délétions de plus de 50 Kpb, sauf 2 parmi les 1226, contenaient chacune au moins 6 des 468 gènes importants en milieu riche avec suivi par passages. C'est en retirant ainsi les délétions de plus de 50 Kpb que nous avons obtenu les 883 délétions mentionnées précédemment qui avaient en moyenne une taille de 3,5 Kpb et qui, combinées, couvraient environ 0,9 Mpb du génome de DGF298. Parmi celles-ci, seulement 16 comportaient certains des 468 gènes classés importants. Par contre, ces dernières délétions étaient surtout retrouvées dans les premiers temps du suivi ou les gènes importants qu'elles contenaient étaient de petite taille (inférieure à 50 pb) et auraient pu être classés ainsi par erreur, à cause de la limite de résolution de la mutagenèse (annexe 2). Également, certaines de ces délétions affectaient des gènes semblant être importants seulement s'ils sont retirés seuls, alors qu'une délétion de leur locus ne causerait pas le même phénotype. Les délétions détectées seraient donc viables, ce qui confirme, malgré le manque de saturation de cette mutagenèse et les différents biais observés, qu'il est possible de réduire davantage le génome de DGF298 sans affecter sa croissance (figure 27).

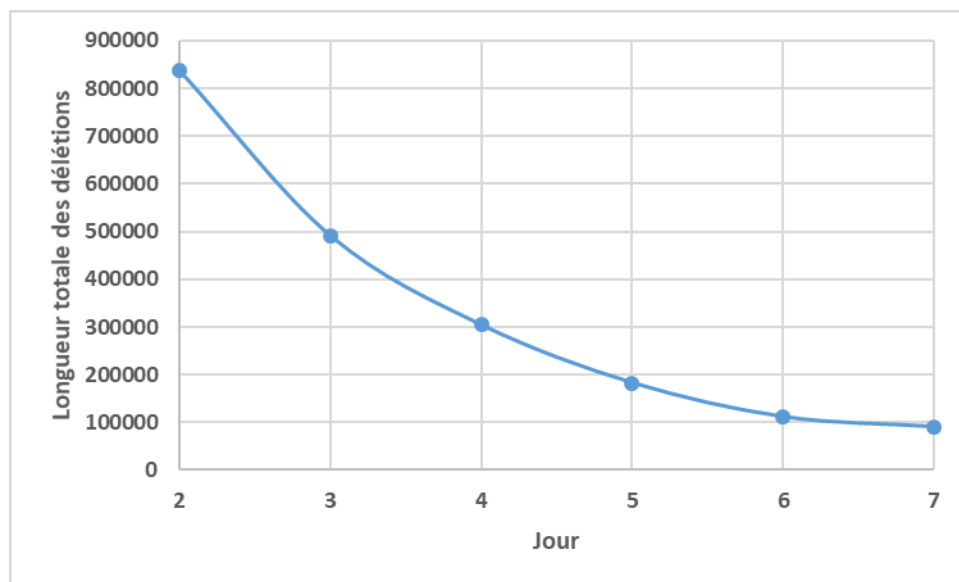


Figure 27. Longueur totale des délétions. Longueur cumulative non redondante des délétions.

Chapitre 4

4. DISCUSSION

Notre étude visait à déterminer si la souche simplifiée d'*Escherichia coli* DGF298 portait encore des gènes facultatifs et pourrait donc potentiellement être réduite davantage. La réduction du génome d'*E. coli* a en effet été entamée depuis déjà plus d'une dizaine d'années par plusieurs groupes en parallèle (Iwadate *et al.*, 2011 ; Hirokawa *et al.*, 2013 ; Karcagi *et al.*, 2016). Une des souches ainsi conçues, DGF298, est celle ayant subi la plus grande réduction, avec un génome obtenu mesurant 3,0 Mpb, tout en conservant une vitesse de croissance similaire à la souche de départ et la capacité de croître dans un milieu minimal. Hirokawa *et al.* (2013) ont suggéré que la taille de DGF298 était près du minimum utile possible pour *E. coli*. Toutefois, plusieurs études soutiennent que le nombre de séquences codantes essentielles dans cette bactérie serait bien moins élevé, jusqu'à 480 (Gerdes *et al.*, 2003 ; Yamamoto *et al.*, 2009 ; Shan *et al.*, 2015). Deux facteurs majeurs sont à prendre en compte pour expliquer cet écart. D'une part, DGF298 a conservé tous les gènes lui permettant de croître normalement, ils sont donc plus nombreux que ceux étant strictement essentiels. D'une autre part, les études concernant l'essentialité des gènes de cette bactérie ont été effectuées avec des souches non réduites. De nombreux gènes peuvent donc paraître facultatifs dû à de la redondance fonctionnelle, mais seraient essentiels chez DGF298 qui aurait perdu, en partie, cette redondance. Pour confirmer que tout le contenu de son génome était important pour une bonne croissance, nous avons effectué des mutagénèses par insertions et une par délétions et avons suivi les mutants dans différentes conditions. Comme expliqué dans l'annexe 3, les mutants d'insertions ont été suivis dans trois conditions distinctes, soit en milieu minimal ou riche avec suivi en turbidostat et en milieu riche avec suivi par passages, mais les données correspondant aux suivis en turbidostat n'ont pas été retenues puisque cette condition impliquerait involontairement une faible diffusion de l'oxygène du milieu vers les cellules, qui devaient alors se servir de leurs voies métaboliques anaérobiques.

4.1 Mutagénèse par insertions

Nous avons utilisé la méthode la plus efficace, à notre connaissance, pour effectuer la mutagénèse par transposons de la souche DGF298. Cette méthode, mettant en jeu la conjugaison d'un plasmide suicide portant un transposon, aurait en effet permis de générer des millions de mutants et d'atteindre une densité d'insertions initiale inégalée, jusqu'à une insertion toutes les deux pb. Cependant, probablement à cause du biais d'insertion de la transposase Tn5, la densité de transposons globale moyenne était de $4,0 \times 10^1$ insertions/Kpb pour les trois temps pris en compte dans un suivi en milieu riche. Ce niveau était suffisant pour déduire l'importance de la plupart des gènes. Nos données soutenaient donc que 2017 gènes n'étaient individuellement pas nécessaires aux cellules pour survivre dans un mélange varié de mutants en compétition. Il est important de souligner l'impact qu'a le fait de suivre non pas une culture simple, mais bien un mélange de mutants. Les gènes requis pour la synthèse d'entérobactine en étaient un exemple. En effet, puisque la méthode impliquait que des millions de mutants différents cohabitent dans une même culture, les cellules incapables de synthétiser de l'entérobactine pouvaient se servir du sidérophore sécrété par les autres mutants. Nous avons donc observé que les gènes impliqués dans la synthèse d'entérobactine (*entA*, *entB*, *entC*, *entD*, *entE*, *entF*) étaient facultatifs pour la croissance en milieu riche avec suivi par passages, alors que ceux impliqués dans son import (*fepA*, *fepB*, *fepC*, *fepD*, *fepG*, *fes*, *tonB*) étaient importants (Raymond *et al.*, 2003). Il serait alors judicieux de combiner ces résultats à une mutagénèse avec cultures simples, comme celle publiée par Baba *et al.* (2006).

La forte compétition exercée sur les mutants a permis d'identifier certains gènes qui n'étaient pas classés importants, mais dont leur interruption pouvait rendre la cellule moins compétitive à long terme. Nous avons ainsi identifié 46 gènes, parmi les 2017 facultatifs, qui pouvaient être importants pour conserver au moins 95% de la vitesse de croissance. Des 1971 séquences codantes semblant avoir peu d'impact sur la croissance, 685 ont aussi été considérées facultatives par plusieurs autres études employant différentes méthodes (Gerdes *et al.*, 2003 ; Kang *et al.*, 2004 ; Hashimoto *et al.*, 2005 ; Baba *et al.*, 2006 ; tableau supplémentaire 1). Dans une perspective de réduction, ce groupe de gènes serait donc le moins risqué, alors qu'il est pratiquement certain que

le retrait de ceux étant classés importants dans les trois échantillons sélectionnés du suivi (468) mènerait à un grand défaut de croissance dans un milieu riche. Le restant des 1971 gènes mentionnés précédemment serait le second groupe en ligne pour la réduction, mais il faudra porter une attention aux cas plus complexes et trompeurs comme celui de la synthèse d'entérobactine. Le retrait des gènes restants possède un niveau de risque croissant selon le nombre d'échantillons supportant leur essentialité. Un gène classé important dans seulement une partie des trois temps pris en compte peut indiquer qu'il a un impact mineur sur la croissance qui est visible seulement dans les derniers temps du suivi ou bien qu'il n'a pas été classé important à cause de la présence de bruit dans au moins un des échantillons. Selon les résultats et en excluant les cas de létalité synthétique, il serait ainsi possible de retirer entre 780 Kpb, selon le groupe le moins risqué confirmé aussi par la littérature (685), et 1922 Kpb, selon toutes les séquences codantes ne semblant pas affecter la croissance (1971), du génome sans perdre la capacité de croître de façon compétitive en milieu riche avec suivi par passages.

L'analyse des fonctions des 1971 gènes facultatifs donne une image globale du choix qui devra être fait lors de la réduction du génome de DGF298, c'est-à-dire de favoriser la simplification de l'organisme, au détriment de son adaptabilité. En effet, beaucoup de ces gènes avaient des fonctions liées à la survie en conditions spécifiques, soit avec une source de carbone autre que le glucose, en absence d'oxygène (anaérobie), avec une carence en certains minéraux, en présence d'antibiotiques ou de métaux toxiques, en milieu acide, à température élevée ou encore sous stress osmotique. Les gènes correspondant à la survie dans ces dernières conditions pourraient donc être retirés en milieu contrôlé, mais la conservation de certains autres étant aussi facultatifs, mais ayant des fonctions importantes en biologie moléculaire, comme la réparation de l'ADN, serait pertinente.

4.2 Mutagénèse par délétions

Bien que les données de mutagénèse par insertions nous indiquent qu'une grande partie des séquences codantes de DGF298 sont facultatives, seul le retrait de celles-ci peut véritablement confirmer qu'elles le sont. En effet, un gène peut, par exemple, sembler facultatif puisqu'il peut contenir des insertions dans la fin de sa séquence, mais sa région N-terminale pourrait, quant à elle, être essentielle. Pour cette raison, nous avons aussi fait une mutagénèse par délétions aléatoires avec une méthode que nous avons développée, impliquant une double transposition et une contre-sélection avec un plasmide conjugatif Cas9. Bien que la méthode ne soit pas encore optimisée, 883 délétions ont été détectées lors du séquençage des temps pris en compte dans le suivi par passages en milieu riche. Cette méthode pourra être optimisée dans le futur pour permettre de créer, sélectionner et détecter davantage de délétions. Outre l'efficacité de l'induction de l'évènement de délétion, trois facteurs principaux seraient à modifier. D'une part, la souche donneuse devrait porter le gène du répresseur cI dans son génome, ce qui éviterait d'avoir à utiliser le plasmide pFG036, qui peut mener à la création de produits de fusion non désirés. Ensuite, le mécanisme d'induction de la seconde transposition, impliquant l'utilisation du chélateur 2,2-bipyridine pour diminuer la concentration en fer, pourrait être modifié, puisqu'il pourrait avoir été la cause du biais initial vers le locus contenant les gènes *cyoABCDE* impliqués dans la chaîne de transport d'électron. En effet, les cellules pourraient bénéficier, lors d'une carence en fer, de la perte de certaines protéines requérant un grand investissement en cet élément (Folsom *et al.*, 2014). Finalement, l'antibiotique utilisé pour sélectionner les mutants ayant survécu à la contre-sélection devrait être changé pour un autre, comme la spectinomycine, puisque la pression de sélection liée à la tétracycline semblait trop élevée. En effet, les mutants ayant le gène *ompR* délété ont rapidement pris le dessus dans la culture et ont ainsi causé une diminution drastique du nombre de délétions uniques présentes. Ce gène est requis pour l'expression des porines *ompF* et *ompC* (Mizuno *et al.*, 1988). Les mutants d'*ompR* étaient donc plus résistants à la tétracycline puisque son entrée par diffusion était limitée (Mortimer and Piddock, 1993).

Dû au fort biais de sélection, il est malheureusement difficile de déterminer si certains mutants de délétion ont été perdus au cours du suivi dû à l'effet de leur délétion spécifiquement ou à cause du trop grand avantage des mutants exempts du gène *ompR*. Nous avons tout de même confirmé qu'il était possible de retirer des régions totalisant 0,9 Mbp du génome de DGF298 avec une mutagénèse par délétions individuelles non optimisée.

Chapitre 5

5. CONCLUSION

L'objectif de notre étude était de déterminer, à l'aide de mutagénèses aléatoires, si le génome de la souche DGF298, mesurant 2 991 126 pb, portait encore des gènes facultatifs à sa croissance et pourrait donc être réduit davantage. Cette étude est donc importante pour l'amélioration des connaissances sur *E. coli*, pour la création d'un organisme minimal à partir de cette bactérie, et, plus globalement, pour l'ingénierie de nouveaux organismes vivants.

Nos résultats de mutagénèses par insertion suggèrent que plus de 73 % des séquences codantes sont individuellement facultatives pour avoir un taux de croissance compétitif en milieu riche. Notre mutagénèse par délétions a confirmé qu'il était possible de retirer des régions totalisant au moins 30 % du génome de DGF298 bien qu'un problème de saturation et une trop forte pression de sélection antibiotique aient rendu plus difficile l'interprétation des résultats de cette mutagénèse. La suite logique de cette étude serait de combiner nos données avec celles d'autres études précédentes pour, dans un premier temps, mettre au point et exécuter un plan de réduction de la souche DGF298 pour obtenir un organisme minimal dérivé d'une protéobactérie. Ensuite, de plus amples modifications pourront être effectuées, ce qui pourra correspondre, cette fois-ci, à un projet de synthèse chimique (Ostrov *et al.*, 2016). Cela pourra être l'occasion de faire des essais téméraires tels que de changer la forme d'*E. coli* et possiblement d'enlever sa paroi et sa membrane externe, ce qui pourrait être pertinent puisque certains croient que la réduction de son génome pourra être limitée non pas directement à cause de son contenu, mais sera plutôt limitée à une taille minimale à cause de la position relative que doivent occuper les gènes essentiels pour leur bon positionnement tridimensionnel, lui-même lié à son type cellulaire (Danchin and Hénaut, 1997 ; Young, 2006 ; Martínez-García and Lorenzo, 2016).

Le travail présenté dans ce mémoire, tout comme celui de plusieurs autres groupes, représente les premiers pas vers une nouvelle ère qui promet de bouleverser profondément notre compréhension du vivant et notre capacité à manipuler les génomes.

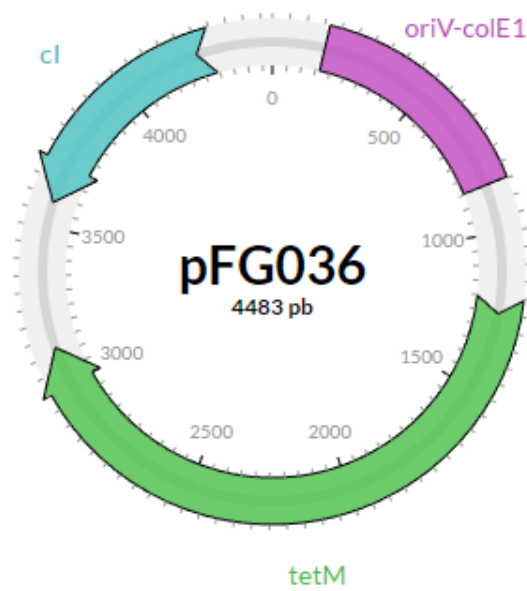
ANNEXE 1

Détails sur la construction des plasmides

Les fragments, obtenus par amplification ou par digestion, requis pour l'assemblage des plasmides et la carte de ces derniers sont affichés dans cette section.

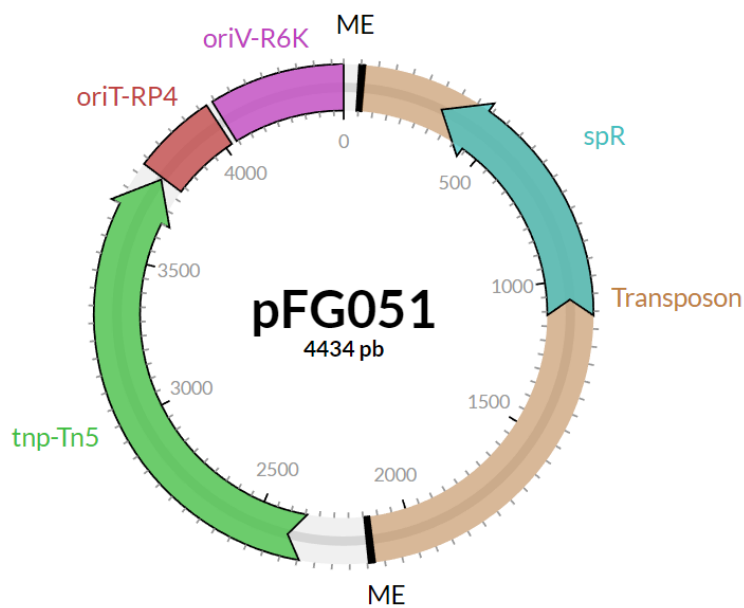
pFG036

Fragment	Provenance	Référence
Mutant thermosensible du répresseur cI	ADN génomique du phage lambda	(Casjens and Hendrix, 2015)
Origine de réplication dérivée du plasmide colE1	pUC19	(Norrande <i>et al.</i> , 1983)
Gène de résistance à la tétracycline	Synthèse (IDT)	-



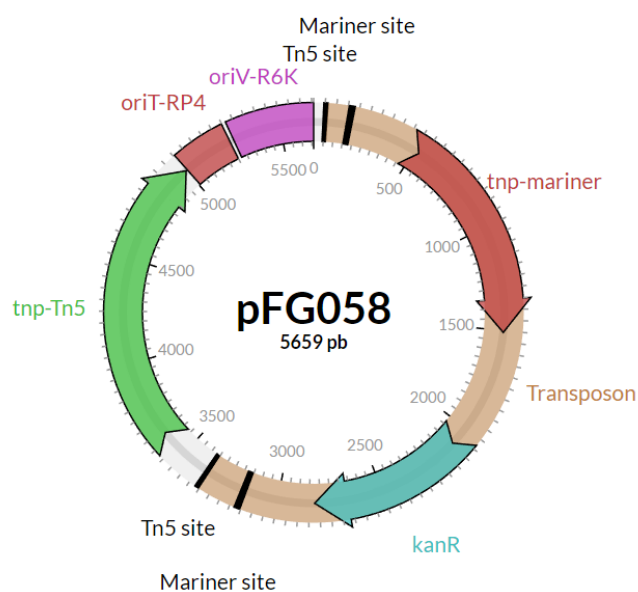
pFG051

Fragment	Provenance	Référence
Promoteur P _L du phage lambda	ADN génomique du phage lambda	(Casjens and Hendrix, 2015)
Gène codant pour la transposase Tn5	pBAM1	(Martinez-Garcia <i>et al.</i> , 2011)
Origine de transfert conjugatif du plasmide RP4 Origine de réplication dérivée du plasmide R6K	pBAM1	(Martinez-Garcia <i>et al.</i> , 2011)
Transposon (séquence contenant un gène de résistance à la spectinomycine et bordée de sites ME)	Synthèse (IDT)	-



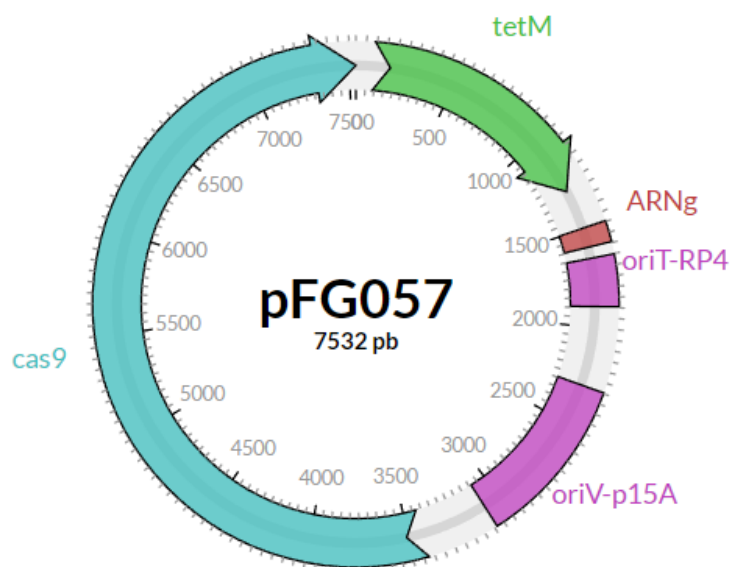
pFG058

Fragment	Provenance	Référence
Promoteur P _L du phage lambda Gène codant pour la transposase Tn5 Origine de transfert conjugatif du plasmide RP4 Origine de réplication dérivée du plasmide R6K	pFG051	Cette étude
Transposon (séquence contenant le gène codant pour la transposase mariner ainsi qu'un gène de résistance à la kanamycine et bordée de sites que reconnaît la transposase mariner et de sites que reconnaît la transposase Tn5)	pSC189 pKD4 pFG051	(Chiang and Rubin, 2002) (Datsenko and Wanner, 2000) Cette étude



pFG057

Fragment	Provenance	Référence
Gène de résistance à la tétracycline	pOSIP-TT	(St-Pierre <i>et al.</i> , 2013)
ARN guide	pFG01	(Grenier <i>et al.</i> , 2015)
Origine de transfert conjugatif du plasmide RP4	pBAM1	(Martinez-Garcia <i>et al.</i> , 2011)
Origine de répllication dérivée du plasmide p15A	pCas9	(Jiang <i>et al.</i> , 2013)
Gène codant pour l'endonucléase Cas9	pCas9	(Jiang <i>et al.</i> , 2013)



ANNEXE 2

Gènes classés importants présents dans les délétions identifiées pour les jours 2 à 7.

Longueur (pb)	Nombre de jours (jours)	Nombre de gènes	Gènes 468 ¹	Explication
5136	1(2)	6	<i>thrC</i>	De la thréonine est présente dans le milieu utilisé. Sa synthèse ne devrait donc pas être essentielle. Les données de la mutagenèse par insertions indiquent cependant qu'elle est importante, possiblement à cause d'un taux d'import insuffisant. Il serait donc possible de trouver des délétions de ce gène dans les premiers temps du suivi.
5817	5(2,3,4,6,7)	7	<i>glxR'</i>	Pseudogène très court (31 pb) probablement classé erronément comme important à cause de la limite de résolution de la mutagenèse.
13122	1(2)	14	<i>trpL</i>	L'opéron <i>trpLEDCBA</i> ne semble pas important selon les données d'insertion. <i>trpL</i> est court (45pb) et probablement classé erronément comme important à cause de la limite de résolution de la mutagenèse.
28938	2(2,5)	31	<i>dicA</i>	<i>dicA</i> réprime l'inhibiteur de division <i>dicB</i> . <i>dicA</i> pourrait donc être important seulement si <i>dicB</i> est présent. Les délétions comprennent les deux gènes.
21964	1(2)	25	<i>dicA</i>	
36824	4(2,3,4,5)	39	<i>dicA</i>	
25191	1(7)	24	<i>dsdA</i>	L'importance de <i>dsdA</i> est difficile à justifier. Il permet de transformer la D-sérine (absente du milieu) en pyruvate et en ammoniac. Les gènes <i>dsdX</i> et <i>dsdC</i> n'ont
5670	3(2,3,4)	5	<i>dsdA</i>	
7391	1(2)	6	<i>dsdA</i>	
2378	1(2)	2	<i>dsdA</i>	

				pas été classés importants. Le premier codant pour un transporteur de D-sérine et le deuxième pour un régulateur des deux autres. <i>dsdA</i> n'a pas été classé important dans un suivi en milieu riche en turbidostat que nous avons fait (voir annexe 3).
196	2(3,4)	1	<i>alaS</i>	Très courte délétion de la fin C-terminale de <i>AlaS</i> (18/2631 pb). Le gène pourrait être encore fonctionnel.
6848	1(2)	8	<i>wzyE</i> , <i>wzxE</i> et <i>rffH</i>	La délétion seule de <i>wzxE</i> est létale dû à l'accumulation de lipides III, excepté si la synthèse de ces lipides est affectée (Rick <i>et al.</i> , 2003). Il est donc possible que de déléter les autres gènes du locus, impliqués dans la synthèse des membranes, en même temps est viable. Cela étant dit, ces délétions étaient présentes uniquement dans le premier échantillon sélectionné et pourraient donc être viables, mais causer un retard considérable de croissance.
607	1(2)	2	<i>wzyE</i>	
4853	1(2)	6	<i>wzyE</i> et <i>wzxE</i>	
6117	2(2,3)	8	<i>atpBFAD</i>	L'ATP synthase n'est pas essentielle, mais son retrait diminue le taux de croissance des cellules (Jensen and Michelsen, 1992). Il serait donc possible de trouver des délétions de ces gènes dans les premiers temps du suivi.
2503	2(2,3)	3	' <i>yjhV</i>	Pseudogène très court (43 pb) probablement classé erronément comme important à cause de la limite de résolution de la mutagenèse.

1. Gènes parmi les 468 classés importants dans les trois temps sélectionnés pour la mutagenèse par insertions

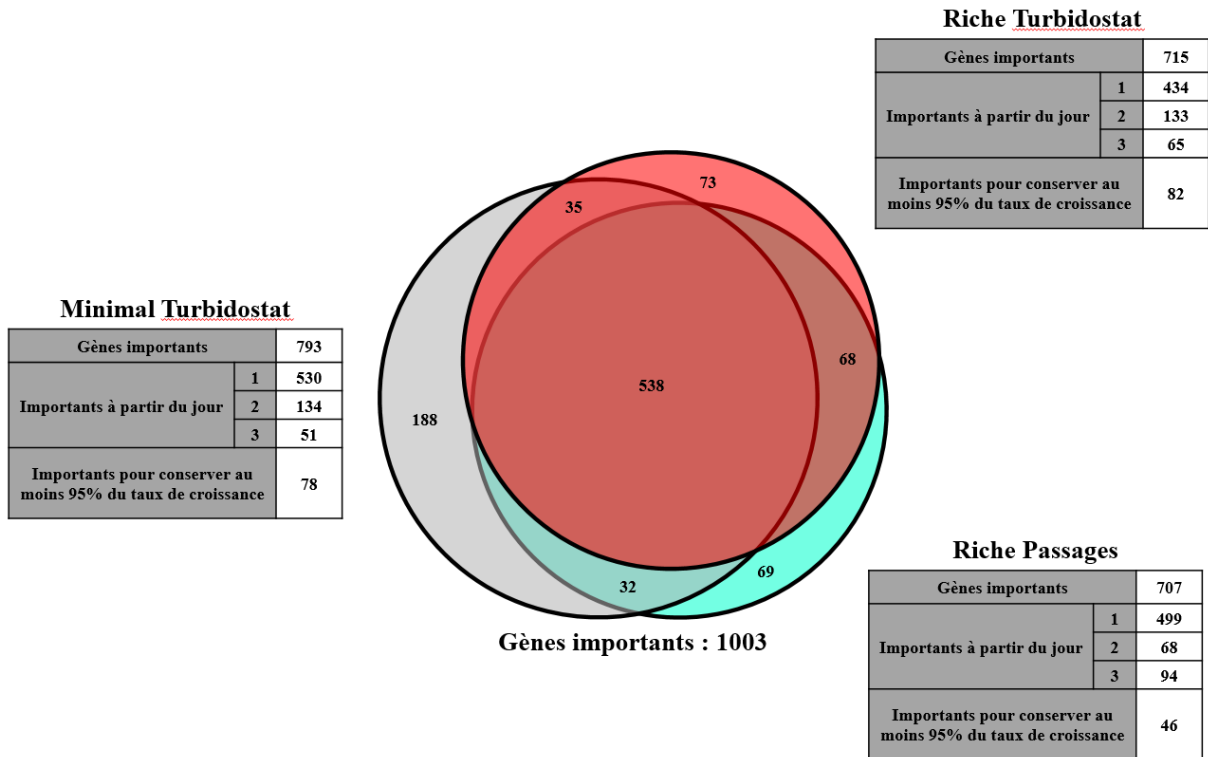
ANNEXE 3

Analyse comparative d'essentialité en milieu riche ou minimal avec suivi en turbidostat ou par passages.

Nous avons fait deux autres suivis, un en milieu minimal et un en milieu riche, en se servant d'un appareil conçu par notre laboratoire permettant de maintenir une culture à une densité optique précise en la diluant automatiquement avec du milieu frais (Matteau *et al.*, 2015). Les mutants suivis de la sorte, obtenus avec 10 gouttelettes de conjugaison, ont été maintenus entre 0.2 et 0.3 de DO_{600} , correspondant à la phase exponentielle, dans 20 mL de milieu avec spectinomycine.

Trois échantillons ont également été sélectionnés pour chacun de ces deux suivis. Les cellules dans ces échantillons ont accompli approximativement 25, 46 et 64 doublements. Les mêmes techniques ont été utilisées pour identifier les gènes importants pour la croissance dans ces conditions. Une comparaison des résultats obtenus pour les trois différents suivis a été faite (figure suivante).

Globalement, 1003 séquences codantes semblaient importantes pour la croissance dans les conditions testées. Une majeure partie (538) de ces gènes était importante dans chacune des conditions. De plus, bien que de nombreux gènes étaient particulièrement importants en milieu minimal (188), le nombre de gènes étant importants pour chacune des conditions, individuellement, était relativement proche (entre 707 et 793), indiquant que certains gènes étaient spécifiquement nécessaires pour la survie dans les autres conditions testées, c'est-à-dire en milieu riche et en suivi en turbidostat ou par passages.



Séquences codantes importantes pour la cellule dans trois conditions distinctes. Diagramme de Venn des séquences codantes classées importantes par TraDIS ou pouvant diminuer, lorsqu’interrompues, le taux de croissance de la cellule à moins de 95 %.

Nous avons ensuite voulu analyser les fonctions des gènes importants spécifiquement en milieu minimal, en milieu riche, en suivi en turbidostat ou en suivi par passages. Durant cette analyse nous avons constaté que les cellules suivies en turbidostat semblaient manquer d’oxygène (malgré l’apport constant en air frais dans le tube de culture), puisque plusieurs gènes (notamment *arcA*, *arcB*, *focA*, *pflA* et *pflB*) requis pour le métabolisme microaérobie ou anaérobie étaient importants spécifiquement pour ce type de suivi (Alexeeva *et al.*, 2000 ; Matteau *et al.*, 2015). Ce phénomène a été observé par Castan and Enfors (2002) et serait dû à de l’ADN, provenant d’une portion des cellules ayant lysé, qui irait entourer les cellules et diminuerait ainsi la diffusion de l’oxygène. Puisque nous n’étions pas intéressés à l’essentialité

en microaérobie, nous avons choisi de conserver uniquement les données du suivi fait par passages.

BIBLIOGRAPHIE

- Andrew S. (2010). FastQC: a quality control tool for high throughput sequence data. Disponible en ligne à : <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2, 2006.0008.
- Barquist, L., Mayho, M., Cummins, C., Cain, A.K., Boinett, C.J., Page, A.J., Langridge, G.C., Quail, M.A., Keane, J.A., and Parkhill, J. (2016). The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics* 32, 1109–1111.
- Blount, Z.D. (2015). The unexhausted potential of *E. coli*. *eLife* 4, e05826.
- Bohn, C., and Boulloc, P. (1998). The *Escherichia coli* *cmlA* gene encodes the multidrug efflux pump Cmr/MdfA and is responsible for isopropyl- β -d-thiogalactopyranoside exclusion and spectinomycin Sensitivity. *J. Bacteriol.* 180, 6072–6075.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Butland, G., Babu, M., Díaz-Mejía, J.J., Bohdana, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A.G., Pogoutse, O., et al. (2008). eSGA: *E. coli* synthetic genetic array analysis. *Nat. Meth.* 5, 789–795.
- Cambray, G., Mutalik, V.K., and Arkin, A.P. (2011). Toward rational design of bacterial genomes. *Curr. Opin. Microbiol.* 14, 624–630.
- Castan, A., and Enfors, S.-O. (2002). Formate accumulation due to DNA release in aerobic cultivations of *Escherichia coli*. *Biotechnol. Bioeng.* 77, 324–328.
- Chiang, S.L., and Rubin, E.J. (2002). Construction of a mariner-based transposon for epitope-tagging and genomic targeting. *Gene* 296, 179–185.
- Danchin, A., and Hénaut, A. (1997). The map of the cell is in the chromosome. *Curr. Opin. Genet. Dev.* 7, 852–854.
- Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *PNAS* 97, 6640–6645.
- Escherich, T. (1988). The intestinal bacteria of the neonate and breast-fed infant. 1884. *Rev. Infect. Dis.* 10, 1220–1225.

- Federowicz, S., Kim, D., Ebrahim, A., Lerman, J., Nagarajan, H., Cho, B., Zengler, K., and Palsson, B. (2014). Determining the Control Circuitry of Redox Metabolism at the Genome-Scale. *PLoS Genet.* *10*, e1004264.
- Ferrières, L., Hémery, G., Nham, T., Guérout, A.-M., Mazel, D., Beloin, C., and Ghigo, J.-M. (2010). Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *J. Bacteriol.* *192*, 6418–6427.
- Fliege, R., Tong, S., Shibata, A., Nickerson, K.W., and Conway, T. (1992). The Entner-Doudoroff pathway in *Escherichia coli* is induced for oxidative glucose metabolism via pyrroloquinoline quinone-dependent glucose dehydrogenase. *Appl. Environ. Microbiol.* *58*, 3826–3829.
- Folsom, J.P., Parker, A.E., and Carlson, R.P. (2014). Physiological and proteomic analysis of *Escherichia coli* iron-limited chemostat growth. *J. Bacteriol.* *196*, 2748–2761.
- Forouhar, F., Kuzin, A., Seetharaman, J., Lee, I., Zhou, W., Abashidze, M., Chen, Y., Yong, W., Janjua, H., Fang, Y., et al. (2007). Functional insights from structural genomics. *J. Struct. Funct. Genomics* *8*, 37–44.
- Gerdes, S.Y., Scholle, M.D., Campbell, J.W., Balázsi, G., Ravasz, E., Daugherty, M.D., Somera, A.L., Kyrpides, N.C., Anderson, I., Gelfand, M.S., et al. (2003). Experimental Determination and System Level Analysis of Essential Genes in *Escherichia coli* MG1655. *J. Bacteriol.* *185*, 5673–5684.
- Goryshin, I.Y., Miller, J.A., Kil, Y.V., Lanzov, V.A., and Reznikoff, W.S. (1998). Tn5/IS50 target recognition. *PNAS* *95*, 10716–10721.
- Goryshin, I.Y., Naumann, T.A., Apodaca, J., and Reznikoff, W.S. (2003). Chromosomal Deletion Formation System Based on Tn5 Double Transposition: Use For Making Minimal Genomes and Essential Gene Analysis. *Genome Res.* *13*, 644–653.
- Green, B., Bouchier, C., Fairhead, C., Craig, N.L., and Cormack, B.P. (2012). Insertion site preference of Mu, Tn5, and Tn7 transposons. *Mob. DNA* *3*, 3.
- Grenier, F., Matteau, D., Baby, V., and Rodrigue, S. (2014). Complete Genome Sequence of *Escherichia coli* BW25113. *Genome Announc.* *2*.
- Grenier, F., Lucier, J.-F., and Rodrigue, S. (2015). Selection and Validation of Spacer Sequences for CRISPR-Cas9 Genome Editing and Transcription Regulation in Bacteria. *Methods Mol. Biol.* *1334*, 233–244.
- Guan, L., Liu, Q., Li, C., and Zhang, Y. (2013). Development of a Fur-dependent and tightly regulated expression system in *Escherichia coli* for toxic protein synthesis. *BMC Biotechnol.* *13*, 25.

- Gyorfy, Z., Draskovits, G., VERNYIK, V., Blattner, F.F., Gaal, T., and Posfai, G. (2015). Engineered ribosomal RNA operon copy-number variants of *E. coli* reveal the evolutionary trade-offs shaping rRNA operon number. *Nucleic Acids Res.* 43, 1783–1794.
- Hashimoto, M., Ichimura, T., Mizoguchi, H., Tanaka, K., Fujimitsu, K., Keyamura, K., Ote, T., Yamakawa, T., Yamazaki, Y., Mori, H., et al. (2005). Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol. Microbiol.* 55, 137–149.
- Hayes, F. (2003). Toxins-Antitoxins: Plasmid Maintenance, Programmed Cell Death, and Cell Cycle Arrest. *Science* 301, 1496–1499.
- Herring, C.D., Glasner, J.D., and Blattner, F.R. (2003). Gene replacement without selection: regulated suppression of amber mutations in *Escherichia coli*. *Gene* 311, 153–163.
- Hirokawa, Y., Kawano, H., Tanaka-Masuda, K., Nakamura, N., Nakagawa, A., Ito, M., Mori, H., Oshima, T., and Ogasawara, N. (2013). Genetic manipulations restored the growth fitness of reduced-genome *Escherichia coli*. *J. Biosci. Bioeng.* 116, 52–58.
- Hutchison, C.A., Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., Ma, L., et al. (2016). Design and synthesis of a minimal bacterial genome. *Science* 351, aad6253.
- Iwadate, Y., Honda, H., Sato, H., Hashimoto, M., and Kato, J. (2011). Oxidative stress sensitivity of engineered *Escherichia coli* cells with a reduced genome. *FEMS Microbiol. Lett.* 322, 25–33.
- Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). CRISPR-assisted editing of bacterial genomes. *Nat. Biotechnol.* 31, 233–239.
- Jullesson, D., David, F., Pfleger, B., and Nielsen, J. (2015). Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. *Biotechnol. Adv.* 33, 1395–1402.
- Junier, I., Martin, O., and Képès, F. (2010). Spatial and Topological Organization of DNA Chains Induced by Gene Co-localization. *PLoS Comput. Biol.* 6, e1000678.
- Kang, Y., Durfee, T., Glasner, J.D., Qiu, Y., Frisch, D., Winterberg, K.M., and Blattner, F.R. (2004). Systematic mutagenesis of the *Escherichia coli* genome. *J. Bacteriol.* 186, 4921–4930.
- Karcagi, I., Draskovits, G., Umenhoffer, K., Fekete, G., Kovács, K., Méhi, O., Balikó, G., Szappanos, B., Györfy, Z., Fehér, T., et al. (2016). Indispensability of Horizontally Transferred Genes and Its Impact on Bacterial Genome Streamlining. *Mol. Biol. Evol.* 33, 1257–1269.
- Kia, A., Gloeckner, C., Osothprarop, T., Gormley, N., Bomati, E., Stephenson, M., Goryshin, I., and He, M.M. (2017). Improved genome sequencing using an engineered transposase. *BMC Biotechnol.* 17, 6.
- Lampe, D.J., Akerley, B.J., Rubin, E.J., Mekalanos, J.J., and Robertson, H.M. (1999). Hyperactive transposase mutants of the Himar1 mariner transposon. *PNAS* 96, 11428–11433.

- Leimbach, A., Hacker, J., and Dobrindt, U. (2013). *E. coli* as an all-rounder: the thin line between commensalism and pathogenicity. *Curr. Top. Microbiol. Immunol.* 358, 3–32.
- Li, W., Yang, B., Liang, S., Wang, Y., Whiteley, C., Cao, Y., and Wang, X. (2008). BLogo: a tool for visualization of bias in biological sequences. *Bioinformatics* 24, 2254–2255.
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. [arXiv:1303.3997v1 \[q-bio.GN\]](https://arxiv.org/abs/1303.3997v1).
- Liu, F., Wang, C., Wu, Z., Zhang, Q., and Liu, P. (2016). A zero-inflated Poisson model for insertion tolerance analysis of genes based on Tn-seq data. *Bioinformatics* 32, 1701–1708.
- Mali, P., Esvelt, K.M., and Church, G.M. (2013). Cas9 as a versatile tool for engineering biology. *Nat. Methods* 10, 957–963.
- Martínez-García, E., and de Lorenzo, V. (2016). The quest for the minimal bacterial genome. *Curr. Opin. Biotechnol.* 42, 216–224.
- Martínez-García, E., Calles, B., Arévalo-Rodríguez, M., and de Lorenzo, V. (2011). pBAM1: an all-synthetic genetic tool for analysis and construction of complex bacterial phenotypes. *BMC Microbiol.* 11, 38.
- Matteau, D., Baby, V., Pelletier, S., and Rodrigue, S. (2015). A Small-Volume, Low-Cost, and Versatile Continuous Culture Device. *PLoS ONE* 10, e0133384.
- Mizuno, T., Kato, M., Jo, Y.L., and Mizushima, S. (1988). Interaction of OmpR, a positive regulator, with the osmoregulated ompC and ompF genes of *Escherichia coli*. Studies with wild-type and mutant OmpR proteins. *J. Biol. Chem.* 263, 1008–1012.
- Mortimer, P.G., and Piddock, L.J. (1993). The accumulation of five antibacterial agents in porin-deficient mutants of *Escherichia coli*. *J. Antimicrob. Chemother.* 32, 195–213.
- Nair, S., and Finkel, S.E. (2004). Dps Protects Cells against Multiple Stresses during Stationary Phase. *J. Bacteriol.* 186, 4192–4198.
- Norlander, J., Kempe, T., and Messing, J. (1983). Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* 26, 101–106.
- Ostrov, N., Landon, M., Guell, M., Kuznetsov, G., Teramoto, J., Cervantes, N., Zhou, M., Singh, K., Napolitano, M.G., Moosburner, M., et al. (2016). Design, synthesis, and testing toward a 57-codon genome. *Science* 353, 819–822.
- Pósfai, G., Kolisnychenko, V., Bereczki, Z., and Blattner, F.R. (1999). Markerless gene replacement in *Escherichia coli* stimulated by a double-strand break in the chromosome. *Nucleic Acids Res.* 27, 4409–4415.
- Pritchard, J.R., Chao, M.C., Abel, S., Davis, B.M., Baranowski, C., Zhang, Y.J., Rubin, E.J., and Waldor, M.K. (2014). ARTIST: High-Resolution Genome-Wide Assessment of Fitness Using Transposon-Insertion Sequencing. *PLoS Genet.* 10, e1004782.

- Raymond, K.N., Dertz, E.A., and Kim, S.S. (2003). Enterobactin: An archetype for microbial iron transport. *PNAS* *100*, 3584–3588.
- Reznikoff, W.S. (2008). Transposon Tn5. *Annu. Rev. Genet.* *42*, 269–286.
- Rice, K.C., and Bayles, K.W. (2008). Molecular control of bacterial death and lysis. *Microbiol. Mol. Biol. Rev.* *72*, 85–109.
- Robertson, H.M., and Lampe, D.J. (1995). Recent horizontal transfer of a mariner transposable element among and between Diptera and Neuroptera. *Mol. Biol. Evol.* *12*, 850–862.
- Rolfe, M.D., Rice, C.J., Lucchini, S., Pin, C., Thompson, A., Cameron, A.D.S., Alston, M., Stringer, M.F., Betts, R.P., Baranyi, J., et al. (2012). Lag Phase Is a Distinct Growth Phase That Prepares Bacteria for Exponential Growth and Involves Transient Metal Accumulation. *J. Bacteriol.* *194*, 686–701.
- Shan, Y., Lazinski, D., Rowe, S., Camilli, A., and Lewis, K. (2015). Genetic Basis of Persister Tolerance to Aminoglycosides in *Escherichia coli*. *mBio* *6*, e00078-15.
- Solaimanpour, S., Sarmiento, F., and Mrázek, J. (2015). Tn-seq explorer: a tool for analysis of high-throughput sequencing data of transposon mutant libraries. *PLoS ONE* *10*, e0126070.
- Tamames, J., González-Moreno, M., Mingorance, J., Valencia, A., and Vicente, M. (2001). Bringing gene order into bacterial shape. *Trends Genet.* *17*, 124–126.
- van Opijnen, T., Bodi, K.L., and Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat. Meth.* *6*, 767–772.
- Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R., and Church, G.M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature* *460*, 894–898.
- Wang, H.H., Kim, H., Cong, L., Jeong, J., Bang, D., and Church, G.M. (2012). Genome-scale promoter engineering by coselection MAGE. *Nat. Methods* *9*, 591–593.
- Yamamoto, N., Nakahigashi, K., Nakamichi, T., Yoshino, M., Takai, Y., Touda, Y., Furubayashi, A., Kinjyo, S., Dose, H., Hasegawa, M., et al. (2009). Update on the Keio collection of *Escherichia coli* single-gene deletion mutants. *Mol. Syst. Biol.* *5*, 335.
- Yang, Z.R., Bullifent, H.L., Moore, K., Paszkiewicz, K., Saint, R.J., Southern, S.J., Champion, O.L., Senior, N.J., Sarkar-Tyson, M., Oyston, P.C.F., et al. (2017). A Noise Trimming and Positional Significance of Transposon Insertion System to Identify Essential Genes in *Yersinia pestis*. *Sci. Rep.* *7*, srep41923.
- Young, K.D. (2006). The Selective Value of Bacterial Shape. *Microbiol. Mol. Biol. Rev.* *70*, 660–703.

Yu, B.J., Sung, B.H., Koob, M.D., Lee, C.H., Lee, J.H., Lee, W.S., Kim, M.S., and Kim, S.C. (2002). Minimization of the *Escherichia coli* genome using a Tn5-targeted Cre/loxP excision system. Nat. Biotech. 20, 1018–1023.

Zomer, A., Burghout, P., Bootsma, H.J., Hermans, P.W.M., and van Hijum, S.A.F.T. (2012). ESSENTIALS: software for rapid analysis of high throughput transposon insertion sequencing data. PLoS ONE 7, e43012.

